

WEST Search History

DATE: Friday, January 03, 2003

<u>Set Name</u> side by side	<u>Query</u>	<u>Hit Count</u>	<u>Set Name</u> result set
<i>DB=USPT,PGPB,EPAB,DWPI; PLUR=YES; OP=OR</i>			
L26	l23 and paclitaxel	1	L26
L25	l23 and paclitaxol	0	L25
L24	L23 and taxol	2	L24
L23	peptide adj internaliz\$	33	L23
<i>DB=USPT,PGPB; PLUR=YES; OP=OR</i>			
L22	peptide adj internaliz\$	32	L22
L21	L20 and taxol	0	L21
L20	L19 and internaliz\$	1	L20
L19	5087616.pn.	1	L19
L18	L17 and internalize	0	L18
L17	5753204.pn.	1	L17
<i>DB=DWPI; PLUR=YES; OP=OR</i>			
L16	WO008800837a2	0	L16
<i>DB=EPAB; PLUR=YES; OP=OR</i>			
L15	WO008800837a2	0	L15
L14	WO008800837a1	0	L14
<i>DB=DWPI; PLUR=YES; OP=OR</i>			
L13	WO008800837a1	0	L13
L12	WO888800837a1	0	L12
L11	WO888800837A2	0	L11
<i>DB=EPAB; PLUR=YES; OP=OR</i>			
L10	L9	0	L10
<i>DB=USPT,PGPB,EPAB,DWPI; PLUR=YES; OP=OR</i>			
L9	WO888800837A2	0	L9
<i>DB=USPT,PGPB; PLUR=YES; OP=OR</i>			
L8	WO888800837A2	0	L8
L7	L6 and internalize	1	L7
L6	L4 and tumor	1	L6
L5	L4 and peptide adj internaliz\$	0	L5
L4	5847121.pn.	1	L4
L3	tumor cell and L2	59243	L3
L2	L1 and internaliz\$	1	L2
L1	5587459.pn.	1	L1

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:27:08 ; Search time 10 seconds
(without alignments)
22.742 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 12
Sequence: 1 TSPLINHGOKL 12

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 117078 seqs, 18951520 residues

Word size : 0

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	100.0	12	US-09-899-376-1	Sequence 1, Appl
2	12	100.0	19	US-09-899-376-3	Sequence 3, Appl
3	12	100.0	19	US-09-899-376-4	Sequence 4, Appl
4	7	58.3	247	US-10-042-417-36	Sequence 36, Appl
5	5	41.7	45	US-09-984-245-157	Sequence 157, App
6	5	41.7	68	US-09-764-869-1142	Sequence 1142, Ap
7	5	41.7	72	US-09-764-877-1500	Sequence 1500, App
8	5	41.7	89	US-09-764-869-920	Sequence 920, App
9	5	41.7	139	US-09-864-761-37407	Sequence 37407, A
10	5	41.7	155	US-09-731-816-5	Sequence 5, Appl
11	5	41.7	155	US-09-886-404-5	Sequence 5, Appl
12	5	41.7	155	US-09-854-280-11	Sequence 11, Appl
13	5	41.7	155	US-09-854-208-11	Sequence 11, Appl
14	5	41.7	158	US-09-764-868-948	Sequence 948, App
15	5	41.7	177	US-08-424-5508-50	Sequence 50, Appl
16	5	41.7	181	US-09-925-300-967	Sequence 967, App
17	5	41.7	194	US-10-164-871-21	Sequence 21, Appl
18	5	41.7	194	US-09-815-242-11500	Sequence 11500, A
19	5	41.7	195	US-10-164-871-2	Sequence 2, Appl

20	5	41.7	212	US-09-782-953-24	Sequence 24, Appl
21	5	41.7	212	US-09-782-953-25	Sequence 25, Appl
22	5	41.7	223	US-10-164-871-22	Sequence 22, Appl
23	5	41.7	235	US-09-828-644-111	Sequence 111, Appl
24	5	41.7	241	US-09-782-953-21	Sequence 21, Appl
25	5	41.7	241	US-09-782-953-22	Sequence 22, Appl
26	5	41.7	300	US-09-925-300-1421	Sequence 1421, Ap
27	5	41.7	382	US-09-815-242-10385	Sequence 10385, A
28	5	41.7	402	US-09-815-242-11350	Sequence 11350, A
29	5	41.7	407	US-09-895-913A-136	Sequence 136, App
30	5	41.7	479	US-09-925-302-527	Sequence 527, App
31	5	41.7	621	US-10-042-417-56	Sequence 56, Appl
32	5	41.7	634	US-09-732-665-10	Sequence 10, Appl
33	5	41.7	635	US-09-732-665-8	Sequence 8, Appl
34	5	41.7	659	US-09-977-260-8	Sequence 8, Appl
35	5	41.7	659	US-09-977-269-8	Sequence 8, Appl
36	5	41.7	717	US-09-817-538-9	Sequence 9, Appl
37	5	41.7	717	US-09-817-538-9	Sequence 9, Appl
38	5	41.7	722	US-09-815-242-10796	Sequence 10796, A
39	5	41.7	917	US-10-001-835-189	Sequence 189, App
40	5	41.7	933	US-10-078-107-5	Sequence 5, Appl
41	5	41.7	933	US-10-077-751-5	Sequence 5, Appl
42	5	41.7	933	US-09-815-242-10095	Sequence 10095, A
43	5	41.7	942	US-09-892-287-4	Sequence 4, Appl
44	5	41.7	942	US-09-969-528-6	Sequence 6, Appl
45	5	41.7	1599	US-10-092-880-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-09-899-376-1
Sequence 1, Application US/09899376
Patent No. US20020102265A1
GENERAL INFORMATION:
APPLICANT: HONG, FRANK D.
TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT
FILE REFERENCE: USCS:64505
CURRENT APPLICATION NUMBER: US/09/899, 376
CURRENT FILING DATE: 2001-07-02
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 12
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-899-376-1

Query Match Best Local Similarity 100.0%; Score 12; DB 10; Length 12;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSPLINHGOKL 12
Db 1 TSPLINHGOKL 12

RESULT 2
US-09-899-376-3
Sequence 3, Application US/09899376
Patent No. US20020102265A1
GENERAL INFORMATION:
APPLICANT: HONG, FRANK D.
APPLICANT: CLAYMAN, GARY
TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT
FILE REFERENCE: USCS:64505

```
; CURRENT APPLICATION NUMBER: US/09/899,376
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-899-376-3

Query Match          100.0%; Score 12; DB 10; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TSPLNHNGQKL 12
        |||
Db      4 TSPLNHNGQKL 15

RESULT 3
US-09-899-376-4
; Sequence 4, Application US/09899376
; Patent No. US20020102265A1
; GENERAL INFORMATION:
; APPLICANT: HONG, FRANK D.
; APPLICANT: CLAYMAN, GARY
; TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT
; FILE REFERENCE: US/09/899,376
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-899-376-4

Query Match          100.0%; Score 12; DB 10; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TSPLNHNGQKL 12
        |||
Db      8 TSPLNHNGQKL 19

RESULT 4
US-10-042-417-36
; Sequence 36, Application US/10042417
; Patent No. US20020123082A1
; GENERAL INFORMATION:
; APPLICANT: Pagano, M.
; TITLE OF INVENTION: METHODS TO IDENTIFY COMPOUNDS USEFUL FOR THE TREATMENT OF
; FILE REFERENCE: 5914-090-999
; CURRENT APPLICATION NUMBER: US/10/042,417
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: 60/260,179
; PRIOR FILING DATE: 2001-01-5
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 36
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
US-10-042-417-36

Query Match          58.3%; Score 7; DB 12; Length 247;
Best Local Similarity 100.0%; Pred. No. 0.51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 IHNGQKL 12
        |||
Db      146 IHNGQKL 152

RESULT 5
US-09-984-245-157
; Sequence 157, Application US/09984245
; Patent No. US20020165374A1
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 87 Human Secreted Proteins
; FILE REFERENCE: P2004P1
; CURRENT APPLICATION NUMBER: US/09/984,245
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/154,707
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: PCT/US98/05311
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: US 60/041,277
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/042,344
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/041,276
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/041,281
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/048,094
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,350
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,188
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,135
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/050,937
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,187
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,099
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,352
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,186
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,069
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,095
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,355
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,160
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,096
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,351
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,154
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/054,804
; PRIOR FILING DATE: 1997-08-05
; PRIOR APPLICATION NUMBER: US 60/056,370
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: US 60/060,862
; PRIOR FILING DATE: 1997-10-02
```



```

; NUMBER OF SEQ ID NOS: 343
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 157
; LENGTH: 45
; TYPE: prt
; ORGANISM: Homo sapiens
;
US-09-984-245-157

```

Query Match	41.7%;	Score 5;	DB 9;	Length 45;
Best Local Similarity	100.0%;	Pred. No. 16;		
Matches	5;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0;

QY	4	LNHN	8
Db	34	LNHN	38

RESULT 6
US-09-764-869-1142
; Sequence 1142, Application US/09764869
; Patent No. US20020061521A1

```

: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC007
: CURRENT APPLICATION NUMBER: US/09/764,869

```

```

: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - refer to PALM or file wrapper
: NUMBER OF SEQ. ID NOS: 2442
: SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 1142
;
; LENGTH: 68
; TYPE: prt
; ORGANISM: Homo sapiens

```

Query Match	41.7%;	Score 5;	DB 10;	Length 68;
Best Local Similarity	100.0%;	Pred. No. 24;		
Matches	5;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;);

QY	2	SPLNI	6
Db	3	SPLNI	7

RESULT 7
US-09-764-877-1500
; Sequence 1500, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:

```

: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC005
: CURRENT APPLICATION NUMBER: US/09/764,877
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - refer to PALM or file wrapper
: NUMBER OF SEQ ID NOS: 4031
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1500

```

```

1  LENGTH: 2
2  TYPE: PRT
3  ORGANISM: Homo sapiens
4  FEATURE:
5  NAME/KEY: SITE
6  LOCATION: (57)
7  OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
8  US-09-764-877-1500

```

Query Match	41.78;	Score 5;	DB 10;	Length 72;
Best Local Similarity	100.08;	Pred. No. 25;		
Matches	5;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

QY 5 NIING 5

Db 39 NIHG 43

RESULT 8
US-09-764-869-920
; Sequence 920, Application US/09764869
; Patent No. US20020061521A1

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007

```

```

:
:
:
: CURRENT APPLICATION NUMBER: US/09/764,869
:
: CURRENT FILING DATE: 2001-01-17
:
: Prior application data removed - refer to PALM or file wrapper
:
: NUMBER OF SEQ ID NOS: 2442
:
: SOFTWARE: PatentIn Ver. 2.0
:

```

```

; SEQ ID NO 92A
;
; LENGTH: 89
; TYPE: prt
; ORGANISM: Homo sapiens

```

```

; NAME/KEY: SITE
; LOCATION: (89)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-869-920

```

Query Match	41.7%;	Score 5;	DB 10;	Length 89;
Best Local Similarity	100.0%;	Pred. No. 30;		
Matches	5;	Conservative 0;	Mismatches 0;	Indels 0;
			Gaps 0;	

QY	3	PLNIH	7
Db	10	PLNIH	14

RESULT 9
US-09-864-761-37407
; Sequence 37407, Application US/09864761
; Patent No. US20020048763A1

```

1 GENERAL INFORMATION:
2
3 APPLICANT: Penn, Sharon G.
4
5 APPLICANT: Rank, David R.
6
7 APPLICANT: Hanzel, David K.
8
9 APPLICANT: Chen, Wensheng
10
11 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
12
13 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

```

```

1 FILE REFERENCE: Aeomic-X-1
2 CURRENT APPLICATION NUMBER: US/09/864,765
3 CURRENT FILING DATE: 2001-05-23
4 PRIOR APPLICATION NUMBER: US 60/180,312
5 PRIOR FILING DATE: 2000-02-04
6 PRIOR APPLICATION NUMBER: US 60/207,456
7 PRIOR FILING DATE: 2000-05-26
8 PRIOR APPLICATION NUMBER: US 09/632,366
9 PRIOR FILING DATE: 2000-08-03

```

1 PRIOR APPLICATION NUMBER: GB 24263.6
2
3 PRIOR FILING DATE: 2000-10-04
4 PRIOR APPLICATION NUMBER: US 60/236,359
5
6 PRIOR FILING DATE: 2000-09-27
7
8 PRIOR APPLICATION NUMBER: PCT/US01/006666
9
10 PRIOR FILING DATE: 2001-01-30
11
12 PRIOR APPLICATION NUMBER: PCT/US01/00667
13
14 PRIOR FILING DATE: 2001-01-30
15
16 PRIOR APPLICATION NUMBER: PCT/US01/00664

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37407
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ALO22395..2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.4
; OTHER INFORMATION: EST_HUMAN HIT: AM837529.1, EVALUATE 5.00e-80
; OTHER INFORMATION: SWISSPROT HIT: Q99593, EVALUATE 7.20e-01
; US-09-864-761-37407

Query Match
Best Local Similarity 41.7%; Score 5; DB 10; Length 139;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLN 5
Db 13 TSPLN 17

RESULT 10
US-09-731-816-5
; Sequence 5, Application US/09731816
; Patent No. US20010023070A1
; GENERAL INFORMATION:
; APPLICANT: Edner, Reinhard
; APPLICANT: Ruben, Steven
; FILE REFERENCE: PF470P1
; CURRENT APPLICATION NUMBER: US/09/731,816
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/169,837
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/320,713
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/087,340
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/099,805
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/131,965
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT US99/11644
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 155
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
```

```

US-09-731-816-5

Query Match
Best Local Similarity 41.7%; Score 5; DB 10; Length 155;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LN1HN 8
Db 49 LN1HN 53

RESULT 11
US-09-886-404-5
; Sequence 5, Application US/09886404
; Patent No. US20020037524A1
; GENERAL INFORMATION:
; APPLICANT: Medlock, Eugene
; APPLICANT: Yen, Richard
; APPLICANT: Silbiger, Scott M.
; APPLICANT: Elliott, Gary S.
; APPLICANT: Nguyen, Hung O.
; APPLICANT: Jing, Shuguan
; TITLE OF INVENTION: IL-17 Like Molecules and Uses Thereof
; FILE REFERENCE: 01017/37128B
; CURRENT APPLICATION NUMBER: US/09/886,404
; CURRENT FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: 09/810,384
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/266,159
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,125
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-886-404-5

Query Match
Best Local Similarity 41.7%; Score 5; DB 10; Length 155;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LN1HN 8
Db 49 LN1HN 53

RESULT 12
US-09-854-280-11
; Sequence 11, Application US/09854280
; Patent No. US20020052027A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Jian
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin
; APPLICANT: Li, Hanzhong
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: IL-17 HOMOLOGOUS POLYPEPTIDES AND THERAPEUTIC USES THEREOF
; FILE REFERENCE: P1381R1C2
; CURRENT APPLICATION NUMBER: US/09/854,280
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 09/311,832
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: US 60/085,579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: US 60/113,621
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 11
; LENGTH: 155
```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-280-11

Query Match
Best Local Similarity 41.7%; Score 5; DB 10; Length 155;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LN1HN 8
    |||||
Db 49 LN1HN 53

RESULT 13
US-09-854-208-11
; Sequence 11, Application US/09854208
; Patent No. US20020106743A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Jian
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin
; APPLICANT: Li, Hanzhong
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: IL-17 HOMOLOGOUS POLYPEPTIDES AND THERAPEUTIC USES
; FILE REFERENCE: P1381-R1
; CURRENT APPLICATION NUMBER: US/09/854,208
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US/09/311,832
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: US 60/085,579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: US 60/113,621
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 11
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-208-11

Query Match
Best Local Similarity 41.7%; Score 5; DB 10; Length 155;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LN1HN 8
    |||||
Db 49 LN1HN 53

RESULT 14
US-09-764-868-948
; Sequence 948, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 948
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (13)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE

```

```

; LOCATION: (148)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (149)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-948

Query Match
Best Local Similarity 41.7%; Score 5; DB 9; Length 158;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SPLN1 6
    |||||
Db 45 SPLN1 49

RESULT 15
US-08-424-550B-50
; Sequence 50, Application US/08424550B
; Patent No. US20020119447A1
; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; APPLICANT: TAMI J. PILOT-MATIAS
; APPLICANT: GEORGE J. DAWSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: SURESH M. DESAI
; APPLICANT: THOMAS P. LEARY
; APPLICANT: ANTHONY SCOTT MUEHROFF
; APPLICANT: JAMES C. ERKER
; APPLICANT: SHERI L. BUIJK
; APPLICANT: ISA K. MUSHARRAF
; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
; NUMBER OF SEQUENCES: 716
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550B
; FILING DATE:
; CLASSIFICATION: 435435
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6565
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-424-550B-50

Query Match
Best Local Similarity 41.7%; Score 5; DB 8; Length 177;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Fri Jan 3 14:42:47 2003

us-09-899-376-1.oli.rapb

Page 6

Qy	8	NGOKL	12
Db	146	NGOKL	150

Search completed: January 3, 2003, 14:29:20
Job time : 10 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:26:48 ; Search time 14 seconds
(without alignments)
25.220 Million cell updates/sec

Title: US-09-899-376-1

Perfect score: 12
Sequence: 1 TSPLEINHCQKL 12

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata1/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata1/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata1/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata1/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	50.0	109	US-09-134-001C-3523	Sequence 3523, Ap
2	6	50.0	457	US-09-134-001C-3235	Sequence 3235, Ap
3	5	41.7	20	US-08-399-696-90	Sequence 90, Appl
4	5	41.7	59	US-08-399-696-101	Sequence 101, Appl
5	5	41.7	155	US-08-432-994A-8	Sequence 8, Appl
6	5	41.7	177	US-08-469-260A-50	Sequence 50, Appl
7	5	41.7	194	US-08-822-264-4	Sequence 4, Appl
8	5	41.7	194	US-09-565-808-21	Sequence 21, Appl
9	5	41.7	195	US-08-960-022-6	Sequence 6, Appl
10	5	41.7	195	US-09-565-808-2	Sequence 2, Appl
11	5	41.7	220	US-08-822-264-1	Sequence 1, Appl
12	5	41.7	223	US-08-822-264-3	Sequence 3, Appl
13	5	41.7	223	US-09-565-808-22	Sequence 22, Appl
14	5	41.7	270	US-08-347-471-4	Sequence 4, Appl
15	5	41.7	277	US-08-347-471-2	Sequence 2, Appl
16	5	41.7	277	US-09-091-952A-26	Sequence 26, Appl
17	5	41.7	277	US-09-091-952A-27	Sequence 27, Appl
18	5	41.7	277	US-09-091-952A-28	Sequence 28, Appl
19	5	41.7	278	US-08-414-926A-4	Sequence 4, Appl
20	5	41.7	278	US-08-926-822-4	Sequence 4, Appl
21	5	41.7	278	US-09-253-682-4	Sequence 4, Appl
22	5	41.7	278	US-09-527-657-4	Sequence 4, Appl
23	5	41.7	285	US-09-091-952A-25	Sequence 25, Appl
24	5	41.7	301	US-08-785-464-1	Sequence 1, Appl
25	5	41.7	349	US-08-605-150A-17	Sequence 17, Appl
26	5	41.7	354	US-09-067-089-2	Sequence 2, Appl
27	5	41.7	356	US-08-270-583-2	Sequence 2, Appl

28	5	41.7	356	1	US-08-783-889A-2	Sequence 2, Appl
29	5	41.7	357	2	US-08-978-182-4	Sequence 4, Appl
30	5	41.7	357	2	US-09-205-681-4	Sequence 4, Appl
31	5	41.7	382	1	US-08-186-833-2	Sequence 2, Appl
32	5	41.7	392	4	US-09-378-088A-90	Sequence 90, Appl
33	5	41.7	539	1	US-08-399-696-122	Sequence 122, Appl
34	5	41.7	597	1	US-08-399-696-102	Sequence 102, Appl
35	5	41.7	626	4	US-08-971-188-10	Sequence 10, Appl
36	5	41.7	626	4	US-09-374-454-21	Sequence 21, Appl
37	5	41.7	652	4	US-09-134-001C-3517	Sequence 3517, Appl
38	5	41.7	659	1	US-08-391-615-2	Sequence 2, Appl
39	5	41.7	659	4	US-08-426-509A-8	Sequence 8, Appl
40	5	41.7	659	5	PCT-US95-05008-8	Sequence 8, Appl
41	5	41.7	739	1	US-07-803-622E-2	Sequence 2, Appl
42	5	41.7	765	1	US-08-425-061-19	Sequence 19, Appl
43	5	41.7	765	2	US-08-825-886-19	Sequence 19, Appl
44	5	41.7	841	4	US-09-413-814-107	Sequence 107, Appl
45	5	41.7	900	1	US-08-425-061-20	Sequence 20, Appl

ALIGNMENTS

```
RESULT 1
US-09-134-001C-3523
: Sequence 3523, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
: FILE REFERENCE: GFC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3523
: LENGTH: 109
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3523

Query Match      50.0%   Score 6;   DB 4;   Length 109;
Best Local Similarity 100.0%;   Pred. No. 7.2;
Matches      6;   Conservative      0;   Mismatches      0;   Indels      0;   Gaps      0;

Oy      1      TSPLENI      6      |||||
Db      40      TSPLENI      45

RESULT 2
US-09-134-001C-3235
: Sequence 3235, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
: FILE REFERENCE: GFC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: PRIOR FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3235
: LENGTH: 467
: TYPE: PRT
```

ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3235

Query Match 50.0%; Score 6; DB 4; Length 467;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NINNGQ 10
|||||
Db 456 NINNGQ 461

RESULT 3
US-08-399-696-90
Sequence 90, Application US/08399696
Patent No. 5756669
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: P53-BINDING POLYPEPTIDES AND
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING SAME
NUMBER OF SEQUENCES: 126
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/399,696
FILING DATE: 02-MAR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/156,671
FILING DATE: 22-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 15522-000710
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 90:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-399-696-90
Query Match 41.7%; Score 5; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPIN 5
|||||
Db 15 TSPIN 19

RESULT 4
US-08-399-696-101
Sequence 101, Application US/08399696
Patent No. 5756669
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: P53-BINDING POLYPEPTIDES AND
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING SAME
NUMBER OF SEQUENCES: 126
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/399,696
FILING DATE: 02-MAR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/156,671
FILING DATE: 22-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 15522-000710
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 101:
SEQUENCE CHARACTERISTICS:
LENGTH: 59 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 19
OTHER INFORMATION: /note= "X is a peptide spacer of at
OTHER INFORMATION: least one amino acid and less than 1000 amino
OTHER INFORMATION: acids; and is a sequence and length selected
OTHER INFORMATION: independently from the spacer located at 40.
FEATURE:
NAME/KEY: Peptide
LOCATION: 40
OTHER INFORMATION: /note= "X is a peptide spacer of at
OTHER INFORMATION: least one amino acid and less than 1000 amino
OTHER INFORMATION: acids; and is a sequence and length selected
OTHER INFORMATION: independently from the spacer located at position 19.
US-08-399-696-101

Query Match 41.7%; Score 5; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPIN 5
|||||
Db 34 TSPIN 38

RESULT 5
US-08-432-994A-8
Sequence 8, Application US/08432994A
Patent No. 6274711
GENERAL INFORMATION:
APPLICANT: Golstein, Pierre
APPLICANT: Rouvier, Eric
APPLICANT: Fossiez, Francois
APPLICANT: Lebecque, Serge J.E.
APPLICANT: Djossou, Odile
APPLICANT: Bancheureau, Jacques
TITLE OF INVENTION: PURIFIED MAMMALIAN CTLA-8 ANTIGENS AND
TITLE OF INVENTION: RELATED REAGENTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: DNAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,994A

FILED DATE: 02-MAY-1995
CLASSIFICATION: 424
PRIORITY APPLICATION: 424
APPLICATION NUMBER: US 08/250,846
FILING DATE: 27-MAY-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/177,747
FILING DATE: 05-JAN-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/077,203
FILING DATE: 14-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: CHING, EDWIN P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0388K3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-852-9196
TELEFAX: 415-496-1200
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 155 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-432-994A-8

Query Match 41.7%; Score 5; DB 4; Length 155;
Best Local Similarity 100.0%; Pred. No. 1,1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LNINH 8
Db 49 LNINH 53

RESULT 6
US-08-469-260A-50
Sequence 50, Application US/08469260A
Patent No. 6451578
GENERAL INFORMATION:
APPLICANT: JOHN N. SIMONS
APPLICANT: TAMU J. PILOT-MATIAS
APPLICANT: GEORGE J. DAWSON
APPLICANT: GEORGE G. SCHLAUDER
APPLICANT: SURESH M. DESAI
APPLICANT: THOMAS P. LEARY
APPLICANT: ANTHONY SCOTT MUEHROFF
APPLICANT: JAMES C. ERKER
APPLICANT: SHERI L. BUIK
APPLICANT: ISA K. MUSHAMMAR
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
NUMBER OF SEQUENCES: 716
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: 100 ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,260A
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/424,550
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: POREMBSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5527.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 177 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-260A-50

Query Match 41.7%; Score 5; DB 4; Length 177;
Best Local Similarity 100.0%; Pred. No. 1,3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 NGOKL 12
Db 146 NGOKL 150

RESULT 7
US-08-822-264-4
Sequence 4, Application US/08822264
Patent No. 6033869
GENERAL INFORMATION:
APPLICANT: GOLL, Surya K.
APPLICANT: Hillman, Jennifer L.
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: NOVEL HUMAN CYTOKINE/STEROID
TITLE OF INVENTION: RECEPTOR PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,264
FILING DATE:
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0233 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1657409
US-08-822-264-4

Query Match 41.7%; Score 5; DB 3; Length 194;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPLN 5
27 TSPLN 31

RESULT 8
US-09-565-808-21
Sequence 21, Application US/09565808
Patent No. 6432674
GENERAL INFORMATION:
APPLICANT: Hirata, Yuichi
TITLE OF INVENTION: STEROID HORMONE BINDING PROTEIN
FILE REFERENCE: 06501-059001
CURRENT APPLICATION NUMBER: US/09/565,808
CURRENT FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: WO/JP98/05010
PRIOR FILING DATE: 1998-11-06
PRIOR APPLICATION NUMBER: JP/9/322376
PRIOR FILING DATE: 1997-11-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 194
TYPE: PRT
ORGANISM: Suidae
US-09-565-808-21

Query Match 41.7%; Score 5; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPLN 5
27 TSPLN 31

RESULT 9
US-08-960-022-6
Sequence 6, Application US/08960022
Patent No. 5976837
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavalie, Edward R.
APPLICANT: Racle, Lisa A.
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vlkl
APPLICANT: Agostino, Michael J.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYPEPTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/960,022
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 195 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-960-022-6

Query Match 41.7%; Score 5; DB 2; Length 195;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPLN 5
27 TSPLN 31

RESULT 10
US-09-565-808-2
Sequence 2, Application US/09565808
Patent No. 6432674
GENERAL INFORMATION:
APPLICANT: Hirata, Yuichi
TITLE OF INVENTION: STEROID HORMONE BINDING PROTEIN
FILE REFERENCE: 06501-059001
CURRENT APPLICATION NUMBER: US/09/565,808
CURRENT FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: WO/JP98/05010
PRIOR FILING DATE: 1998-11-06
PRIOR APPLICATION NUMBER: JP/9/322376
PRIOR FILING DATE: 1997-11-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 195
TYPE: PRT
ORGANISM: Homo sapiens
US-09-565-808-2

Query Match 41.7%; Score 5; DB 4; Length 195;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPLN 5
27 TSPLN 31

RESULT 11
US-08-822-264-1
Sequence 1, Application US/08822264
Patent No. 6033869
GENERAL INFORMATION:
APPLICANT: Goli, Surya K.
APPLICANT: Hillman, Jennifer L.
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: NOVEL HUMAN CYTOKINE/STEROID
TITLE OF INVENTION: RECEPTOR PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,264
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0233 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 220 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: CONUTUT101
CLONE: 2504333
US-08-822-264-1

Query Match 41.7%; Score 5; DB 3; Length 220;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPIN 5
11111
DB 27 TSPIN 31

RESULT 12
US-08-822-264-3
Sequence 3, Application US/08822264
Patent No. 6033869
GENERAL INFORMATION:
APPLICANT: Goll, Surya K.
APPLICANT: Hillman, Jennifer L.
APPLICANT: Murtry, Lynn E.
TITLE OF INVENTION: NOVEL HUMAN CYTOKINE/STEROID
TITLE OF INVENTION: RECEPTOR PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,264
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0233 US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 158818
US-08-822-264-3

Query Match 41.7%; Score 5; DB 3; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPIN 5
11111
DB 27 TSPIN 31

RESULT 13
US-09-565-808-22
Sequence 22, Application US/09565808
Patent No. 6432674
GENERAL INFORMATION:
APPLICANT: Hirata, Yuichi
TITLE OF INVENTION: STEROID HORMONE BINDING PROTEIN
FILE REFERENCE: 06501-059001
CURRENT APPLICATION NUMBER: US/09/565,808
PRIOR FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: WO/99/05010
PRIOR FILING DATE: 1998-11-06
PRIOR APPLICATION NUMBER: JP/9/322376
PRIOR FILING DATE: 1997-11-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 22
LENGTH: 223
TYPE: PRT
ORGANISM: Rattus rattus
US-09-565-808-22

Query Match 41.7%; Score 5; DB 4; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPIN 5
11111
DB 27 TSPIN 31

RESULT 14
US-08-347-471-4
Sequence 4, Application US/08347471
Patent No. 5710015
GENERAL INFORMATION:
APPLICANT: McAllister, George
APPLICANT: Whiting, Paul J.
TITLE OF INVENTION: CDNA CLONING OF INOSITOL MONOPHOSPHATASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Robert J. No. 5710015ch
STREET: P.O. Box 2000, 126 E. Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/347,471
FILING DATE: 08-DECEMBER-94
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 5710015th, Robert J.
REGISTRATION NUMBER: 27,366
REFERENCE/DOCKET NUMBER: T-1160
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-7262
TELEX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-347-471-4

Query/Match 41.7%; Score 5; DB 1; Length 270;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 NGOKL 12
|||||
Db 135 NGOKL 139

RESULT 15
US-08-347-471-2
Sequence 2, Application US/08347471
Patent No. 5710015
GENERAL INFORMATION:
APPLICANT: McAllister, George
APPLICANT: Whiting, Paul J.
TITLE OF INVENTION: CDNA CLONING OF INOSITOL MONOPHOSPHATASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Robert J. No. 5710015th
STREET: P.O. Box 2000, 126 E. Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/347,471
FILING DATE: 08-DECEMBER-94
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 5710015th, Robert J.
REGISTRATION NUMBER: 27,366
REFERENCE/DOCKET NUMBER: T-1160
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-7262
TELEX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 277 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-347-471-2

Query/Match 41.7%; Score 5; DB 1; Length 277;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 NGOKL 12
|||||
Db 142 NGOKL 146

Search completed: January 3, 2003, 14:29:03
Job time : 15 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:26:13 ; Search time 28 Seconds
(without alignments)
88.306 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 12
Sequence: 1 TSPININGOKL 12

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhnc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvivirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	58.3	173	4	09UKC5	09UKC5	homo sapien
2	58.3	188	2	093MT1	093MT1	photohabitu
3	58.3	321	4	096ER2	096ER2	homo sapien
4	58.3	367	11	091V19	091V19	mus musculus
5	58.3	415	11	09JIE4	09JIE4	mus musculus
6	58.3	471	4	09H0V2	09H0V2	homo sapien
7	58.3	480	11	09DC63	09DC63	mus musculus
8	58.3	480	11	09D2J5	09D2J5	mus musculus
9	50.0	158	10	08VXX3	08VXX3	arabidopsis
10	50.0	169	2	051528	051528	pseudomonas
11	50.0	169	16	09HVN0	09HVN0	pseudomonas
12	50.0	393	2	09JR00	09JR00	actinobacil
13	50.0	420	5	017057	017057	caenorhabdi
14	50.0	706	12	089581	089581	african swi
15	50.0	706	12	089926	089926	african swi
16	50.0	1050	3	000666	000666	emericeila

17	50.0	1690	5	077165	077165	maestigmoeb
18	41.7	33	4	08WMC1	08WMC1	homo sapien
19	41.7	48	8	034187	034187	crocodylus
20	41.7	61	16	08YMT5	08YMT5	anabaena sp
21	41.7	63	5	09VLS4	09VLS4	drosohilla
22	41.7	63	16	09PNP4	09PNP4	campylobact
23	41.7	68	4	096GH5	096GH5	homo sapien
24	41.7	69	5	09NMK8	09NMK8	leishmania
25	41.7	69	16	092H95	092H95	ricketsia
26	41.7	76	12	098519	098519	parametium
27	41.7	80	16	P74809	P74809	synecocyst
28	41.7	88	2	09A1G1	09A1G1	carsonella
29	41.7	91	17	08TLK8	08TLK8	methanosarc
30	41.7	100	11	09WVN3	09WVN3	mus musculus
31	41.7	100	17	058804	058804	pyrococcus
32	41.7	104	11	070206	070206	mus musculus
33	41.7	108	11	09D1V8	09D1V8	mus musculus
34	41.7	110	12	082463	082463	hantavirus
35	41.7	116	5	08T0J4	08T0J4	drosohilla
36	41.7	117	4	09POL8	09POL8	homo sapien
37	41.7	119	8	09ME61	09ME61	beta vulgar
38	41.7	122	4	09NM99	09NM99	homo sapien
39	41.7	123	4	09PIN7	09PIN7	homo sapien
40	41.7	124	16	09A473	09A473	caulobacter
41	41.7	127	2	093IC1	093IC1	staphylococ
42	41.7	127	3	09UVI2	09UVI2	candida tro
43	41.7	128	10	004349	004349	arabidopsis
44	41.7	128	16	092B26	092B26	listeria in
45	41.7	129	10	09LUM8	09LUM8	arabidopsis

ALIGNMENTS

RESULT 1

ID Q9UKC5 PRELIMINARY; PRT; 173 AA.
AC Q9UKC5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE F-box protein Fbx3 (Fragment).
GN Fbx3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20003060; PubMed=10531035;
RA Cenciarelli C., Chlaur D.S., Guardavaccaro D., Parks W., Vidal M.,
RA Pagano M.,
RT "Identification of a family of human F-box proteins.";
RL Curr. Biol. 9:1177-1179(1999).
DR EMBL; AF174595; AAF04516.1; -;
DR InterPro; IPR001810; F-box.
DR Pfam; PF00646; F-box; 1.
DR SMART; SM00256; FBOX; 1.
DR PROSITE; PSS0181; FBOX; 1.
FT NON_TER
SQ SEQUENCE 173 AA; 20087 MW; DA01B22A57DAA4E CRC64;

Query Match 58.3%; Score 7; DB 4; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGKL 12
|||||
Db 146 IHNGKL 152

RESULT 2
Q93MT1

ID Q93MT1 PRELIMINARY: PRT; 188 AA.
 AC Q93MT1;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE M1E.
 GN M1E.
 OS Photorhabdus luminescens (Xenorhabdus luminescens).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Photorhabdus
 OX NCBI_TaxID=29488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K122;
 RA Meslier L., Pimenta A., Blight M.:
 "Cloning and analysis of a mannose resistant fimbrial operon from the
 entomopathogenic bacterium, Photorhabdus luminescens K122.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF96083; AAR82428.1; -
 DR InterPro: IPR000259; Fimbril.
 DR Pfam: PF00419; Fimbril; 1.
 DR PRINTS: PRO1613; FIMBRIALPAPF.
 SQ SEQUENCE 188 AA; 20383 MW; CE6DE89285C40C8 CRC64;

Query Match 58.3%; Score 7; DB 2; Length 188;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGOKL 12
 |||||
 DB 138 IHNGOKL 144

RESULT 3
 ID Q96ER2 PRELIMINARY: PRT; 321 AA.
 AC Q96ER2;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Unknown (protein for MGC:21259).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS, AND EMBRYONAL CARCINOMA;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC012012; AAH12012.1; -
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; Zf_C2H2; 7.
 DR ProDom; PD000003; Znf_C2H2; 6.
 DR PROSITE; PS50805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_6.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
 DR DNA-binding; Zinc-finger.
 KW SM-binding; Zinc-finger.
 SQ SEQUENCE 321 AA; 36369 MW; 62BE342C8B7067D9 CRC64;

Query Match 58.3%; Score 7; DB 4; Length 321;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGOKL 12
 |||||
 DB 313 IHNGOKL 319

RESULT 4
 Q91VI9

ID Q91VI9 PRELIMINARY: PRT; 367 AA.
 AC Q91VI9;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Hypothetical 41.9 kDa protein.
 GN FBX03.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BREAST TUMOR;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC010212; AAH10212.1; -
 DR MGD; MGI:1929084; FBX03.
 KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 41861 MW; FD1FDB6DBC46EA9F CRC64;

Query Match 58.3%; Score 7; DB 11; Length 367;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGOKL 12
 |||||
 DB 35 IHNGOKL 41

RESULT 5
 ID Q9JIE4 PRELIMINARY: PRT; 415 AA.
 AC Q9JIE4;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE F-box protein FBA.
 GN FBX03 OR FBA.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20399565; PubMed=10945168;
 RA Ilyin G.P., Riialand M., Pigeon C., Gugen-Guillouzo C.;
 RT "cDNA Cloning and Expression Analysis of New Members of the Mammalian
 F-box Protein Family.";
 RL Genomics 67:40-47(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ilyin G.P.;
 RT "F-box protein FBA.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF233226; AAF67156.1; -
 DR MGD; MGI:1929084; FBX03.
 DR InterPro: IPR001810; F-box.
 DR Pfam; PF00646; F-box; 1.
 DR SMART; SM00256; FBOX; 1.
 DR PROSITE; PS50181; FBOX; 1.
 SQ SEQUENCE 415 AA; 47453 MW; 2CF5277BD0C4DE2A CRC64;

Query Match 58.3%; Score 7; DB 11; Length 415;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGOKL 12
 |||||
 DB 148 IHNGOKL 154

RESULT 6

```

Q9H0V2
ID Q9H0V2 PRELIMINARY; PRT; 471 AA.
AC Q9H0V2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 54.6 kDa protein.
GN DKF2P564B092.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Well B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansoerge W., Boecker M., Bloecker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.W., Oltenselder B., Obermaier B., Tampe J., Heubner D.,
RA Wambolt R., Korn B., Klein M., Poustka A.;
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL Genome Res. 11:422-435(2001).
DR EMBL; AL136625; CAB66560.1; -.
DR InterPro; IPR001810; F-box.
DR Pfam; PF00646; F-box; 1.
DR SMART; SM00256; FBOX; 1.
DR PROSITE; PSS0181; FBOX; 1.
KW Hypothetical protein.
SQ SEQUENCE 471 AA; 54572 MW; CCFD0DDC8A884410 CRC64;

Query Match 58.3%; Score 7; DB 4; Length 471;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGOKL 12
Db 148 IHNGOKL 154

RESULT 7
Q9DC63 PRELIMINARY; PRT; 480 AA.
ID Q9DC63;
AC Q9DC63;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 120002G09Rik protein.
GN FBX03 OR 120002G09Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LUNG;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

```

```

RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK004544; BAB23360.1; -.
DR MGD; MGI:1929084; Fbxo3.
DR InterPro; IPR001810; F-box.
DR Pfam; PF00646; F-box; 1.
DR SMART; SM00256; FBOX; 1.
DR PROSITE; PSS0181; FBOX; 1.
SQ SEQUENCE 480 AA; 55227 MW; 9AD9565076A67787 CRC64;

Query Match 58.3%; Score 7; DB 11; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGOKL 12
Db 148 IHNGOKL 154

RESULT 8
Q9D2J5 PRELIMINARY; PRT; 480 AA.
ID Q9D2J5;
AC Q9D2J5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 120002G09Rik protein.
GN FBX03 OR 120002G09Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK019550; BAB31793.1; -.
DR MGD; MGI:1929084; Fbxo3.
DR InterPro; IPR001810; F-box.
DR Pfam; PF00646; F-box; 1.
DR SMART; SM00256; FBOX; 1.
DR PROSITE; PSS0181; FBOX; 1.
SQ SEQUENCE 480 AA; 55245 MW; ACA21584B8EF5482 CRC64;

Query Match 58.3%; Score 7; DB 11; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGOKL 12
Db 148 IHNGOKL 154

```

DB 148 IHHGOKL 154

RESULT 9

08VXX3 PRELIMINARY; PRT; 158 AA.

ID 08VXX3

AC 08VXX3

DT 01-MAR-2002 (TEMBLrel. 20, Created)

DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)

DE Hypothetical 18.2 kDa protein.

GN A15G58740.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucotsids II; Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;

OK NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,

RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Ouach H.L.,

RA Tang C., Tortolmi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,

RA Carinuci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,

RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,

RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,

RA Sakurai T., Satou M., Seki M., Shim P., Southwick A., Shinzaki K.,

RA Davis R.W., Ecker J.R., Theologis A.;

RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY074384; AAL67080.1; -;

DR InterPro: IPR005307; Nuclear_move.

DR Pfam: PF03593; Nuclear_move; 1.

DR Hypothetical protein.

OK SEQUENCE 158 AA; 18211 MW; 8764B24FCC8F1BF3 CRC64;

Query Match

Best Local Similarity 50.0%; Score 6; DB 10; Length 158;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHHGOK 11

DB 14 IHHGOK 19

RESULT 10

051528

ID 051528

AC 051528

DT 01-NOV-1996 (TEMBLrel. 01, Created)

DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)

DT 01-NOV-1996 (TEMBLrel. 08, Last annotation update)

DE Fimbrial protein FIMT.

GN FIMT.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

NCBI_TaxID=287;

OK NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN-PA01.

RE MEDLINE=96272255; PubMed=8682785;

RA Alm R.A., Matick J.S.;

RT Identification of two genes with prepilin-like leader sequences

involved in type 4 fimbrial biogenesis in Pseudomonas aeruginosa.;

RL J. Bacteriol. 178:3809-3817(1996).

CC -1; FUNCTION: INVOLVED IN TYPE 4 FIMBRIAL BIOGENESIS.

DR EMBL; L48934; AAB39270.1; -;

KW Fimbrial.

OK SEQUENCE 169 AA; 18958 MW; EB08764E1FF2AAEC CRC64;

Query Match

Best Local Similarity 50.0%; Score 6; DB 2; Length 169;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNIHNG 9

DB 123 LNIHNG 128

RESULT 11

09HVN0

ID 09HVN0

AC 09HVN0

DT 01-MAR-2001 (TEMBLrel. 16, Created)

DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)

DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)

DE Type 4 fimbrial biogenesis protein FIMT.

GN FIMT OR PA4549.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

NCBI_TaxID=287;

OK NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=ATCC 15692 / PA01.

RE MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V., an

RT opportunistic pathogen.;

RL Nature 406:959-964(2000).

DR EMBL; AE004869; AAG07937.1; -;

OK COMPLETE proteome.

OK SEQUENCE 169 AA; 18889 MW; EB08764E1FF2AAEC CRC64;

Query Match

Best Local Similarity 50.0%; Score 6; DB 16; Length 169;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNIHNG 9

DB 123 LNIHNG 128

RESULT 12

09JRO0

ID 09JRO0

AC 09JRO0

DT 01-OCT-2000 (TEMBLrel. 15, Created)

DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)

DE Putative glycosyltransferase.

OS Actinobacillus actinomycetemcomitans (Haemophilus

actinomycetemcomitans).

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Actinobacillus.

NCBI_TaxID=714;

OK NCBI_TaxID=714;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN-C01000;

RE MEDLINE=21391794; PubMed=11500407;

RA Kaplan J.B., Perry M.B., Maclean L.L., Furgang D., Wilson M.E.,

RA Fine D.H.;

RT Structural and Genetic Analyses of O Polysaccharide from

Actinobacillus actinomycetemcomitans Serotype I.;

RL Infect. Immun. 69:5375-5384(2001).

DR EMBL; AF213680; AAF25887.2; -;

KW Transferase.

OK SEQUENCE 393 AA; 46619 MW; C625EFOEA7C14C5A CRC64;

Query Match

Best Local Similarity 50.0%; Score 6; DB 2; Length 393;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LNIHNG 9
Db 221 LNIHNG 226

RESULT 13

017057

ID 017057 PRELIMINARY; PRT; 420 AA.

AC 017057;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-OCT-2001 (TREMBLrel. 18, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE Hypothetical 46.4 kDa protein.

GN C24A1.2

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE=99069613; PubMed=9851916;

RA None;

RT "Genome sequence of the nematode C. elegans: a platform for

RT investigating biology. The C. elegans Sequencing Consortium.";

RT Science 282:2012-2018(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Connell M.;

RT "The sequence of C. elegans cosmid C24A1.1";

RT Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Waterston R.;

RT "Direct Submission.";

RT Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.

DR HSSP; P14921; 25TT.

DR InterPro; IPR000418; Ets.

DR Pfam; PF00178; Ets; 1.

DR PRINTS; PR00454; ETSDOMAIN.

DR SMART; SM00413; ETS; 1.

DR PROSITE; PS00345; ETS_DOMAIN_1; 1.

DR PROSITE; PS00346; ETS_DOMAIN_2; 1.

DR PROSITE; PS50061; ETS_DOMAIN_3; 1.

KW Hypothetical protein.

SQ SEQUENCE 420 AA; 46383 MW; AAIDF31254563459 CRC64;

Query Match 50.0%; Score 6; DB 5; Length 420;
Best Local Similarity 100.0%; Pred. No. 33;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPLNT 6
Db 21 TSPLNT 26

RESULT 14

089581

ID 089581 PRELIMINARY; PRT; 706 AA.

AC 089581;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE HELICASE.

GN 0706L

OS African swine fever virus (ASFV).

OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.

OX NCBI_TaxID=10497;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BA71V;

RX MEDLINE=96036500; PubMed=7483270;

RX Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,

RA Rodriguez J.F., Vinuela E.;

RT "Immune protection conferred by the baculovirus-related glycoprotein

RT of rhogoto virus (Orthomyxoviridae).";

RT Virology 208:249-278(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BA71V;

RX MEDLINE=94233765; PubMed=8178480;

RX la Vega I., Gonzalez A., Blasco R., Calvo V., Vinuela E.;

RT "Nucleotide sequence and variability of the inverted terminal

RT repetitions of African swine fever virus DNA.";

RT Virology 201:152-156(1994).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BA71V;

RX MEDLINE=90219205; PubMed=2325203;

RX Gonzalez A., Calvo V., Almazan F., Almendral J.M., Ramirez J.C.,

RA La Vega I., Blasco R., Vinuela E.;

RT "Multigene families in African swine fever virus: family 360.";

RT J. Virol. 64:2073-2081(1990).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN-BA71V;

RX MEDLINE=90219204; PubMed=2325202;

RX Almendral J.M., Almazan F., Blasco R., Vinuela E.;

RT "Multigene families in African swine fever virus: family 110.";

RT J. Virol. 64:2064-2072(1990).

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN-BA71V;

RX MEDLINE=91134988; PubMed=1994575;

RA Camacho A., Vinuela E.;

RT "Protein p22 of African swine fever virus: an early structural protein

RT that is incorporated into the membrane of infected cells.";

RT Virology 181:251-257(1991).

RN [6]

RP SEQUENCE FROM N.A.

RC STRAIN-BA71V;

RX Almazan F., Murguía J.R., Rodriguez J.M., la Vega I., Vinuela E.;

RL Submitted (APR-1995) to the EMBL/Genbank/DBJ databases.

RN [7]

RP SEQUENCE FROM N.A.

RC STRAIN-BA71V;

RX MEDLINE=94187118; PubMed=8139051;

RX Rodriguez J.M., Yanez R.J., Pan R., Rodriguez J.F., Salas M.L.,

RA Vinuela E.;

RT "Multigene families in African swine fever virus: family 505.";

RT J. Virol. 68:2746-2751(1994).

RN [8]

RP SEQUENCE FROM N.A.

RC STRAIN-BA71V;

RX MEDLINE=93346971; PubMed=8393914;

RX Yanez R.J., Rodriguez J.M., Rodriguez J.F., Salas M.L., Vinuela E.;

RT "African swine fever virus thymidylate kinase gene: sequence and

RT transcriptional mapping.";

RT J. Gen. Virol. 74:1633-1638(1993).

RN [9]

RP SEQUENCE FROM N.A.

RC STRAIN-BA71V;

RX MEDLINE=94065656; PubMed=8245848;

RX Alami A., Angulo A., Vinuela E.;

RT "Mapping and sequence of the gene encoding the African swine fever

RT virion protein of M(1) 11500.";

RT J. Gen. Virol. 74:2317-2324(1993).

RN [10]

RP SEQUENCE FROM N.A.

RC STRAIN-BA71V;

RA MEDLINE-93277388; PubMed-8503790;
 RA Munoz M., Freije J.M., Salas M.L., Vinuela E., Lopez-Otin C.;
 RT "Structure and expression in E. coli of the gene coding for protein
 RT p10 of African swine fever virus.";
 RL Arch. Virol. 130:93-107(1993).
 RN [11]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-90357780; PubMed-2389555;
 RA Blasco R., Lopez-Otin C., Munoz M., Bockamp E.O., Simon-Mateo C.,
 RA Vinuela E.;
 RT "Sequence and evolutionary relationships of African swine fever virus
 RT thymidine kinase.";
 RL Virology 178:301-304(1990).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-93281390; PubMed-8506138;
 RA Yanez R.J., Bourisnell M., Nogal M.L., Yuste L., Vinuela E.;
 RT "African swine fever virus encodes two genes which share significant
 RT homology with the two largest subunits of DNA-dependent RNA
 RT polymerases.";
 RL Nucleic Acids Res. 21:2423-2427(1993).
 RN [13]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-93353606; PubMed-8102411;
 RA Rodriguez J.M., Yanez R.J., Almazan F., Vinuela E., Rodriguez J.F.;
 RT "African swine fever virus encodes a CD2 homolog responsible for the
 RT adhesion of erythrocytes to infected cells.";
 RL J. Virol. 67:512-520(1993).
 RN [14]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-94085774; PubMed-8262374;
 RA Yanez R.J., Rodriguez J.M., Bourisnell M., Rodriguez J.F., Vinuela E.;
 RT "Two putative African swine fever virus helicases similar to yeast
 RT 'DEH' pre-mRNA processing proteins and vaccinia virus ATPases D1L
 RT and D6R.";
 RL Gene 134:161-174(1993).
 RN [15]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-90223993; PubMed-2327074;
 RA Lopez-Otin C., Freije J.M., Parra F., Mendez E., Vinuela E.;
 RT "Mapping and sequence of the gene coding for protein p72, the major
 RT capsid protein of African swine fever virus.";
 RL Virology 175:477-484(1990).
 RN [16]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-94123986; PubMed-8293992;
 RA Rodriguez J.M., Yanez R.J., Rodriguez J.F., Vinuela E., Salas M.L.;
 RT "The DNA polymerase-encoding gene of African swine fever virus:
 RT sequence and transcriptional analysis.";
 RL Gene 136:103-110(1993).
 RN [17]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-93327788; PubMed-8335009;
 RA Simon-Mateo C., Andres G., Vinuela E.;
 RT "Polyprotein processing in African swine fever virus: a novel gene
 RT expression strategy for a DNA virus.";
 RL EMBO J. 12:2977-2987(1993).
 RN [18]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-93233210; PubMed-8474154;
 RA Prados F.J., Vinuela E., Alcamí A.;
 RT "Sequence and characterization of the major early phosphoprotein p32
 RT of African swine fever virus.";
 RL J. Virol. 67:2475-2485(1993).
 RN [19]

RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-92260660; PubMed-1583732;
 RA Alcamí A., Angulo A., Lopez-Otin C., Munoz M., Freije J.M.,
 RA Carrascosa A.L., Vinuela E.;
 RT "Amino acid sequence and structural properties of protein p12, an
 RT African swine fever virus attachment protein.";
 RL J. Virol. 66:3860-3868(1992).
 RN [20]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-93174976; PubMed-8438592;
 RA Yanez R.J., Vinuela E.;
 RT "African swine fever virus encodes a DNA ligase.";
 RL Virology 193:531-536(1993).
 RN [21]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-93174941; PubMed-8382399;
 RA Pena L., Yanez R.J., Revilla Y., Vinuela E., Salas M.L.;
 RT "African swine fever virus guanylyltransferase.";
 RL Virology 193:319-328(1993).
 RN [22]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-93174941; PubMed-8382399;
 RA Pena L., Yanez R.J., Revilla Y., Vinuela E., Salas M.L.;
 RT "African swine fever virus guanylyltransferase.";
 RL Virology 193:319-328(1993).
 RN [23]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-95159428; PubMed-7856088;
 RA Simon-Mateo C., Freije J.M., Andres G., Lopez-Otin C., Vinuela E.;
 RT "Mapping and sequence of the gene encoding protein p17, a major
 RT African swine fever virus structural protein.";
 RL Virology 206:1140-1144(1995).
 RN [24]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-92263807; PubMed-1316688;
 RA Garcia-Beato R., Freije J.M., Lopez-Otin C., Blasco R., Vinuela E.,
 RA Salas M.L.;
 RT "A gene homologous to topoisomerase II in African swine fever virus.";
 RL Virology 188:938-947(1992).
 RN [25]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-94091056; PubMed-8266720;
 RA Freije J.M., Lain S., Vinuela E., Lopez-Otin C.;
 RT "Nucleotide sequence of a nucleoside triphosphate phosphohydrolase
 RT gene from African swine fever virus.";
 RT gene from African swine fever virus.";

Query Match 50.0%; Score 6; DB 12; Length 706;
 Best Local Similarity 100.0%; Fred No. 49;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SPLINH 7
 |||||
 Db 485 SPLINH 490

RESULT 15
 Q89926 PRELIMINARY: PRT: 706 AA.
 ID 089926;
 AC 089926;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE HELICASE.
 GN J10L.
 OS African swine fever virus (ASFV).
 OC viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
 OX NCBI_TaxID=10497;
 RX SEQUENCE FROM N.A.
 RC STRAIN-MALAWI L1L20 /1;
 RX MEDLINE-94014996; PubMed-8409937;
 RA Vydelingum S., Baylis S.A., Bristow C., Smith G.L., Dixon L.K.;
 RT "Duplicated genes within the variable right end of the genome of a


```
RT pathogenic isolate of African swine fever virus.";  
RN J. Gen. Virol. 74:2125-2130(1993).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MALAWI LIL20 /1;  
RX MEDLINE=94292916; PubMed=8021596;  
RA Dixon L.K., Twigg S.R.F., Baylis S.A., Vydelingum S., Bristow C.,  
Hammond J.M., Smith G.L.;  
RT "Nucleotide sequence of a 55 kbp region from the right end of the  
genome of a pathogenic African swine fever virus isolate (Malawi  
LIL20/1).";  
RL J. Gen. Virol. 7:1655-1684(1994).  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93393435; PubMed=8397501;  
RA Roberts P.C., Lu Z., Kulish G.F., Rock D.L.;  
RT "Three adjacent genes of African swine fever virus with similarity to  
essential poxvirus genes.";  
RL Arch. Virol. 132:331-342(1993).  
[4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MALAWI;  
RX MEDLINE=93389456; PubMed=8376971;  
RA Baylis S.A., Twigg S.R.F., Vydelingum S., Dixon L.K., Smith G.L.;  
RT "Three African swine fever virus genes encoding proteins with homology  
to putative helicases of vaccinia virus.";  
RL J. Gen. Virol. 74:1969-1974(1993).  
EMBL: X71982; CAA50829.1; -;  
DR EMBL: M88275; AAA03222.1; -;  
DR EMBL: X72952; CAA51457.1; -;  
DR InterPro: IPR001650; Helicase_C.  
DR Pfam: PF00271; Helicase_C: 1.  
DR SMART: SM00490; HelicC: 1.  
KW ATP-binding; Helicase.  
SQ SEQUENCE 706 AA; 80498 MW; 0D22C089781FMAA46 CRC64;  
  
Query Match 50.0%; Score 6; DB 12; Length 706;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 2 SPLNII 7  
IIIIII  
Db 485 SPLNII 490
```

Search completed: January 3, 2003, 14:28:20
Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:25:12 ; Search time 11 Seconds
(without alignments)
45.247 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 12
Sequence: 1 TSPLEINHCOKL 12

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.3	321	1	Z177_HUMAN	Q13360 homo sapien
2	58.3	471	1	FBX3_HUMAN	Q9UK99 homo sapien
3	50.0	294	1	YOGI_BACSU	P46340 bacillus su
4	50.0	391	1	Y534_METJA	Q57954 methanococc
5	50.0	1273	1	YAR2_SCHPO	Q10135 schizosacch
6	41.7	111	1	YTC8_YEAST	P47063 saccharomyc
7	41.7	113	1	SENA_APLCA	P29233 aplysia cal
8	41.7	124	1	SPH2_SULSO	Q9UW98 sulfolobus
9	41.7	126	1	FX18_CANMA	Q00680 candida mal
10	41.7	126	1	FX18_CANTR	P22009 candida tro
11	41.7	137	1	CPC_CUGSA	P29602 cucumis sat
12	41.7	155	1	IL17_HUMAN	Q16552 homo sapien
13	41.7	188	1	UBCX_PICAN	Q60013 pichia angu
14	41.7	193	1	PGC1_PIG	Q95230 sus scrofa
15	41.7	194	1	MINC_HELPJ	Q9ZM51 helicobacte
16	41.7	194	1	PGC1_HUMAN	Q00264 homo sapien
17	41.7	194	1	PGC1_MOUSE	Q55022 mus musculu
18	41.7	194	1	PGC1_RAT	P70580 rattus norv
19	41.7	195	1	MINC_HELPY	Q56693 helicobacte
20	41.7	208	1	YF4F_SCHPO	Q14080 schizosacch
21	41.7	228	1	TE2D_PLAFA	P32086 plasmodium
22	41.7	241	1	CCP3_HUMAN	Q9UKA8 homo sapien
23	41.7	251	1	KDKA_VIRCH	Q9KVB9 vibrio chol
24	41.7	265	1	YG2T_YEAST	P53262 saccharomyc
25	41.7	266	1	KLR8_MOUSE	Q60682 mus musculu
26	41.7	273	1	ILVE_THEMA	P74921 thermotoga
27	41.7	277	1	MYOP_BOVIN	P20455 bos taurus
28	41.7	277	1	MYOP_HUMAN	P29218 homo sapien
29	41.7	277	1	MYOP_MOUSE	Q55023 mus musculu
30	41.7	277	1	MYOP_RAT	P97697 rattus norv
31	41.7	283	1	VP40_MABVM	P35260 marburg vir
32	41.7	285	1	MYOP_XENLA	P39219 xenopus lae
33	41.7	289	1	TF_CAVPO	Q9J1U8 cavia porce

34	5	41.7	292	1	TF_RABIT	P24055 oryctolagus
35	5	41.7	294	1	CGG1_MOUSE	P51945 mus musculu
36	5	41.7	294	1	CGG1_RAT	P39950 rattus norv
37	5	41.7	299	1	RPBD_SALTY	P26392 salmonella
38	5	41.7	303	1	VP40_MABYP	Q03040 marburg vir
39	5	41.7	304	1	CAH4_VARY	P33065 variola vir
40	5	41.7	306	1	YFV9_SCHPO	Q9P796 schizosacch
41	5	41.7	307	1	DNAB_SYNP3	P73097 synecocyst
42	5	41.7	337	1	DEGA_BACSU	P37947 bacillus su
43	5	41.7	338	1	YAXL_RHISN	P55704 rhizobium s
44	5	41.7	341	1	RECA_LACIA	Q39486 lactococcus
45	5	41.7	344	1	GUN4_TRIPE	Q14405 trichoderma

ALIGNMENTS

RESULT 1	ID	2177_HUMAN	STANDARD:	PRT:	321 AA.
AC	Q13360:				
DT	01-NOV-1997	(Rel. 35, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Zinc finger protein 177.				
GN	ZNF177.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID:9606;				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE:96299641; PubMed:8661005;				
RA	Baban S., Freeman J.D., Mager D.L.;				
RT	"Transcripts from a novel human KRAB zinc finger gene contain spliced				
RT	Alu and endogenous retroviral segments.";				
RL	Genomics 33:463-472(1996).				
CC	-1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.				
CC	-1- SUBCELLULAR LOCATION: Nuclear (Probable).				
CC	-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.				
CC	-1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: U37263; AB09749.1; -.				
DR	HSSP: P08047; ISP2.				
DR	GeneW: HGNC:12966; ZNF177.				
DR	MIM: 601276; -.				
DR	InterPro: IPR001909; KRAB.				
DR	InterPro: IPR000822; ZNF_C2H2.				
DR	Pfam: PF00096; ZF-C2H2; 7.				
DR	Pfam: PF01352; KRAB; 1.				
DR	PRINTS: PR00048; ZINCFINGER.				
DR	ProDom: PD000003; Znf_C2H2; 7.				
DR	SMART: SM00349; KRAB; 1.				
DR	SMART: SM00355; ZNF_C2H2; 7.				
DR	PROSITE: PS50805; KRAB; 1.				
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.				
DR	PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.				
KW	Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;				
KW	Nuclear protein; Repeat.				
FT	DOMAIN	14	84		KRAB.
FT	DOMAIN	124	314		ZINC_FINGERS.
FT	ZN_FING	124	146		C2H2-TYPE.
FT	ZN_FING	152	174		C2H2-TYPE.
FT	ZN_FING	180	202		C2H2-TYPE.

```

FT ZN_FING 208 230 C2H2-TYPE.
FT ZN_FING 236 258 C2H2-TYPE.
FT ZN_FING 264 286 C2H2-TYPE.
FT ZN_FING 292 314 C2H2-TYPE.
SQ SEQUENCE 321 AA; 36473 MW; E3258606C292DA17 CRC64;

Query Match 58.3%; Score 7; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 IHNGOKL 12
Db 313 IHNGOKL 319

RESULT 2
FBX3_HUMAN
ID FBX3_HUMAN STANDARD; PRT; 471 AA.
AC Q9UK99; Q9NWX2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE F-box only protein 3.
DE FBX3 OR FBX3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN 1[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta.
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "Mito human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN 2[2]
RP SEQUENCE OF 6-415 FROM N.A.
RX MEDLINE=20003061; PubMed=10531037;
RA Winston J.T., Koepf D.M., Zhu C., Elledge S.J., Harper J.W.;
RT "A family of mammalian F-box proteins.";
RL Curr. Biol. 9:1180-1182(1999).

CC -!- FUNCTION: Probably recognizes and binds to some phosphorylated
CC proteins and promotes their ubiquitination and degradation.
CC -!- SUBUNIT: PART OF A SCF (SKP1-CULLIN-F-BOX) PROTEIN LIGASE COMPLEX
CC (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 F-BOX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AK001943; BA91991.1; -
DR EMBL: AF176702; AA03702.1; -
DR Genew: HGNC:13582; FBX03.
DR InterPro: IPR001810; F-box.
DR Pfam: PF00646; F-box; 1.
DR SMART: SM00256; FBOX; 1.
DR PROSITE: PS50181; FBOX; 1.
DR Ubl conjugation.
KW F-BOX.
FT DOMAIN 419 452 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 453 456 POLY-ARG.
FT DOMAIN 463 466 POLY-ARG.
FT CONFLICT 164 164 T -> A (IN REF. 2).
FT CONFLICT 414 415 EM -> VS (IN REF. 2).
SQ SEQUENCE 471 AA; 54590 MW; F7AA88193E14E67E CRC64;

```

```

Query Match 58.3%; Score 7; DB 1; Length 471;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 IHNGOKL 12
Db 148 IHNGOKL 154

RESULT 3
YOG1_BACSU
ID YOG1_BACSU STANDARD; PRT; 294 AA.
AC P46340;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable ABC transporter permease protein yog1.
DE YOG1.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=1423;
RN 1[1]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / JH642;
RA Takemaru K.-I., Mizuno M., Kobayashi Y.;
RT "A Bacillus subtilis gene cluster similar to the Escherichia coli
RT phosphate-specific transport (pst) operon: evidence for a tandemly
RT arranged pstB gene.";
RL Microbiology 142:2017-2020(1996).

RN 2[2]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / JH642;
RA MEDLINE=97124195; PubMed=869508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the Bacillus subtilis genome containing the skin element and many
RL sporulation genes.";
RN 3[3]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolydin A., Borchert S.,
RA Boursier L., Boursier L., Brans A., Braun M., Bristell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.D., Daniel R.A.,
RA Denizot F., Devigne K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Flitz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gilm S.Y., Glaser P., Goffeau A., Goldlithy E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Hengst A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Meisel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Portoullis S., Prescott A.M.,
RA Prescan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadie Y.,
RA Sato T., Scanlan E., Schleich S., Schreier R., Scottone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzengger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus

```

```

RT subllis."
RL Nature 390:249-256(1997).
CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC YOGGIHUX, PROBABLY RESPONSIBLE FOR THE TRANSLOCATION OF THE
CC SUBSTRATE ACROSS THE MEMBRANE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM PERMEASE FAMILY, CYSTW SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; D58414; BAA09583.1; -
CC EMBL; D84432; BAA12512.1; -
CC EMBL; Z59116; CAB14427.1; -
CC Subllist; BG11377; YGGI.
CC InterPro: IPR000515; BPD_transp.
CC Pfam: PF00528; BPD_transp.1.
CC PROSITE: PS00402; BPD_TRANS_P1N_MEMBR.1.
CC Hypothetical protein; Transmembrane; Complete proteome.
CC TRANSMEM 14 34 POTENTIAL.
CC TRANSMEM 66 86 POTENTIAL.
CC TRANSMEM 106 124 POTENTIAL.
CC TRANSMEM 132 149 POTENTIAL.
CC TRANSMEM 190 210 POTENTIAL.
CC TRANSMEM 260 280 POTENTIAL.
CC SEQUENCE 294 AA; 31462 MW; 8E3F56C2DF365591 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 294;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPINI 6
Db 230 TSPINI 235

RESULT 4
Y534_METJA STANDARD; PRT; 391 AA.
ID Y534_METJA STANDARD; PRT; 391 AA.
AC 057954;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0534.
GN MJ0534.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcidococaceae; Methanococcidococcus.
ON NCBI_TaxID=2190;
RX NCBI_TaxID=2190;
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Overhage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Ruhnmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Borodovsky M.,
RA Colton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: TO M.JANNASCHII MJ0732.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U67502; AAB98525.1; -
CC HSSP; Q9F0J6; IE5D.
CC TIGR; MJ0534; -
CC InterPro: IPR001279; Bactamase-like.
CC InterPro: IPR001226; Flavodoxin.
CC Pfam; PF00258; Flavodoxin.1.
CC Pfam; PF00753; Lactamase_B.1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 391 AA; 45532 MW; C08636FC7C18D516 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 391;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPINI 6
Db 283 TSPINI 288

RESULT 5
YAR2_SCHPO STANDARD; PRT; 1273 AA.
ID YAR2_SCHPO STANDARD; PRT; 1273 AA.
AC Q10135;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C23E2.02 in chromosome I.
GN SPAC23E2.02.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
ON NCBI_TaxID=4896;
RX NCBI_TaxID=4896;
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hildalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voiclaert G., Aert R., Robben J., Grymposiez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesti D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Galfreau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Motier S.,
RA Gallbrecht F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez S., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Useery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: TO M.JANNASCHII MJ0732.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; Z68887; CAA93114.1; -.
 DR InterPro; IPR002937; Amino-oxidase.
 DR InterPro; IPR000910; HMG_12_box.
 DR Pfam; PF01593; Amino-oxidase; 1.
 DR SMART; SM00398; HMG; 1.
 DR Hypothetical protein; Transmembrane.
 FT TRANSMEM 1028 1048 POTENTIAL.
 SQ SEQUENCE 1273 AA; 142489 MW; 3EEFF6433DF3C0B6 CRC64;

Query Match 50.0%; Score 6; DB 1; Length 1273;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SPLN1H 7
 Db 790 SPLN1H 795

RESULT 6
 Y0C8_YEAST STANDARD; PRT; 111 AA.
 AC P47062;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 12.9 kDa protein in MMD2-RNR2 intergenic region.
 GN YJL028W OR J1267.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN 119
 RP SEQUENCE FROM N.A.
 RA Pohl T.M., Aljinovic G.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; Z49303; CAA89319.1; -.
 DR SGD; S0003565; YJL028W.
 DR Hypothetical protein; Transmembrane.
 FT TRANSMEM 52 72 POTENTIAL.
 FT TRANSMEM 52 72 POTENTIAL.
 SQ SEQUENCE 111 AA; 12897 MW; 610024D82C821589 CRC64;

Query Match 41.7%; Score 5; DB 1; Length 111;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSP1N 5
 Db 24 TSP1N 28

RESULT 7
 SENA_APLCA STANDARD; PRT; 113 AA.
 ID SENA_APLCA
 AC P29233;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Sensorin A precursor.
 GN PSC1.
 OS Aplysia californica (California sea hare).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidae;
 OC Aplysiidae; Aplysia.
 OX NCBI_TaxID=6500;
 RN 119
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 33-54.
 RC TISSUE=pleural sensory cells;
 RX MEDLINE=91227915; Pubmed=1840700.
 RA Brunet J.-F., Shapiro E., Foster S.A., Kandel E.R., Iino Y.;
 RT Identification of a peptide specific for Aplysia sensory neurons by
 RT PCR-based differential screening.";
 RL Science 252:856-859(1991).
 CC -1- FUNCTION: MAY FUNCTION AS AN INHIBITORY COTRANSMITTER ACTING IN
 CC CONJUNCTION WITH THE FAST EXCITATORY TRANSMITTER RELEASED BY
 CC SENSORY NEURONS. THE PEPTIDE SELECTIVELY INHIBITS CERTAIN
 CC POSTSYNAPTIC CELLS PROBABLY BY MEANS OF SENSORIN A RELEASE.
 CC -1- SUBCELLULAR LOCATION: THROUGHOUT THE NEURONAL CELLS (CELL BODY,
 CC AXON & PRESYNAPTIC TERMINALS).
 CC -1- TISSUE SPECIFICITY: SEEMS TO BE SPECIFIC TO THE MECHANOSENSORY
 CC NEURONS OF THE CENTRAL NERVOUS SYSTEM.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; X56770; CAA40089.1; -.
 DR PIR; S23653; S23653.
 DR Signal; Neuropeptide; Cleavage on pair of basic residues.
 FT SIGNAL 1 32
 FT PEPTIDE 33 54 POTENTIAL.
 FT PEPTIDE 46 54 PEPTIDE B.
 FT PROPEP 58 113 SENSORIN A.
 FT MOD_RES 54 54
 SQ SEQUENCE 113 AA; 12711 MW; 74350F5154B49ELA CRC64;

Query Match 41.7%; Score 5; DB 1; Length 113;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSP1N 5
 Db 7 TSP1N 11

RESULT 8
 SPH2_SULSO STANDARD; PRT; 124 AA.
 ID SPH2_SULSO
 AC Q9UW78;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE S-adenosylmethionine decarboxylase proenzyme 2 (EC 4.1.1.50) (AdometPC
 DE 2) (Sampc 2) [containing: S-adenosylmethionine decarboxylase 2 beta
 DE chain; S-adenosylmethionine decarboxylase 2 alpha chain].
 GN SPH2 OR S500585.
 OS Sulfolobus solifarius.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=2287;
 RN 119
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; Pubmed=11427726;
 RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aayez M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Etrauso G., Fletcher C., Gordon P.M.K.,


```

RT oxidation."
RL EBF. J. Biochem. 190:107-112(1990).
CC -1- FUNCTION: ITS PHYSIOLOGICAL ROLE IS UNKNOWN, BUT POSSESSES A
CC NONSPECIFIC LIPID-TRANSFER ACTIVITY, DESPITE THE ABSENCE OF A
CC CYSTEINE RESIDUE THOUGHT TO BE ESSENTIAL FOR THE ACTIVITY OF ITS
CC MAMMALIAN COUNTERPARTS.
CC -1- PATHWAY: BETA-OXIDATION OF LONG-CHAIN FATTY ACIDS (PROBABLE).
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Peroxisomal.
CC -1- INDUCTION: BY ALKANES OR FATTY ACIDS; REPRESSED BY GLUCOSE.
CC -1- SIMILARITY: HIGH, TO MAMMALIAN LIVER NONSPECIFIC LIPID-
CC TRANSFER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X53633; CAA37682.1; -.
DR EMBL: M24440; AAA56863.1; -.
DR PIR: JS0155; JS0155.
DR PIR: S11327; S11327.
DR InterPro: IPR003033; SCP2.
DR Pfam: PF02036; SCP2; 1.
DR KMW Lipid-binding; Peroxisome.
FT INIT_MET 0
FT DOMAIN 32 40 HYDROPHOBIC.
FT 42 51 HYDROPHILIC.
SQ SEQUENCE 126 AA; 13675 MW; F2DD99F50CBF87F0 CRC64;

Query Match 41.7%; Score 5; DB 1; Length 126;
Best local Similarity 100.0%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NGORKL 12
DB 91 NGORKL 95

RESULT 11
CPC_CUCSA STANDARD: PRT: 137 AA.
ID CPC_CUCSA
AC P29602;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cucurbit peelings cupredoxin (CPC) (Stellacyanin).
OS Cucumis sativus (Cucumber).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotids I; Cucurbitales; Cucurbitaceae; Cucumis.
OX NCBI_TaxID=3659;
RN [1]
RP SEQUENCE.
RC TISSUE=Peelings; PubMed=1468551;
RX MEDLINE=93106154;
RA Mann K., Schaefer W., Thoenes U., Messerschmidt A., Mehrhahn Z.,
RA Nalbandyan R.;
RT "The amino acid sequence of a type I copper protein with an unusual
RT serine- and hydroxyproline-rich C-terminal domain isolated from
RT cucumber peelings."
RL FEBS Lett. 314:220-223(1992).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX MEDLINE=97084802; PubMed=8931136;
RA Hart P.J., Nersisyan A.M., Herrmann R.G., Nalbandyan R.M.,
RA Valentin J.S., Eisenberg D.;
RT "A missing link in cupredoxins: crystal structure of cucumber
RT stellacyanin at 1.6-A resolution."
RL Protein Sci. 5:2175-2183(1996).

```

```

CC -1- SIMILARITY: CONTAINS 1 PLASTOCYANIN-LIKE DOMAIN.
DR PIR: S26511; SSRV.
DR PDB: 1JER; 12-FEB-97.
DR InterPro: IPR000923; BlueCu_1.
DR InterPro: IPR003245; P1cyanin_1like.
DR Pfam: PF02296; Cu_Bind_1like; 1.
DR ProDom: PD003122; P1cyanin_1like; 1.
DR PROSITE: PS00196; COPPER_BLUE; 1.
KW Electron transport; Copper; Glycoprotein; Hydroxylation; 3D-structure.
FT DOMAIN 1 137
FT MOD_RES 1 137
FT METAL 46 46 PYRROLIDONE CARBOXYLIC ACID.
FT METAL 89 89 COPPER (BY SIMILARITY).
FT METAL 94 94 COPPER (BY SIMILARITY).
FT METAL 99 94 COPPER (BY SIMILARITY).
FT DISULFID 60 95 BY SIMILARITY.
FT CARBOHYD 109 109 N-LINKED (GLCNAC...).
FT MOD_RES 115 115 HYDROXYLATION (PARTIAL).
FT MOD_RES 116 116 HYDROXYLATION.
FT MOD_RES 117 117 HYDROXYLATION.
FT MOD_RES 121 121 HYDROXYLATION.
FT MOD_RES 122 122 HYDROXYLATION.
FT MOD_RES 128 128 HYDROXYLATION.
FT MOD_RES 129 129 HYDROXYLATION.
FT MOD_RES 132 132 HYDROXYLATION (PARTIAL).
FT MOD_RES 133 133 HYDROXYLATION.
FT MOD_RES 134 134 HYDROXYLATION.
FT MOD_RES 136 136 HYDROXYLATION.
SQ SEQUENCE 137 AA; 14747 MW; F15FBC5B34A8CB CRC64;

Query Match 41.7%; Score 5; DB 1; Length 137;
Best local Similarity 100.0%; Pred. No. 39;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NGORKL 12
DB 97 NGORKL 101

RESULT 12
IL17_HUMAN STANDARD: PRT: 155 AA.
ID IL17_HUMAN
AC Q16552;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Interleukin-17 precursor (IL-17) (IL-17A) (Cytotoxic T lymphocyte-
DE associated antigen 8) (CTLA-8).
GN IL17 OR IL17A OR CTLA8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96281911; PubMed=8676080;
RA Fossiez F., Djossou O., Chomaret P., Flores-Romo L., Alt-Yahia S.,
RA Meat C., Plin J.-J., Garrone P., Garcia E., Selsel N., Blanchard D.,
RA Gallard C., Das Mahapatra B., Rouvier E., Golstein P., Bancheau J.,
RA Lebecque S.;
RT "T cell interleukin-17 induces stromal cells to produce
RT proinflammatory and hematopoietic cytokines."
RL J. Exp. Med. 183:2593-2603(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=T-cell;
RX MEDLINE=96094436; PubMed=7499828;
RA Yao Z., Painter S.L., Fanslow W.C., Ulrich D., Macduff B.M.,
RA Spriggs M.K., Armitage R.J.;
RT "Human IL-17: a novel cytokine derived from T cells."
RL J. Immunol. 155:5483-5486(1995).
CC -1- FUNCTION: INDUCES STROMAL CELLS TO PRODUCE PROINFLAMMATORY AND
CC HEMATOPOIETIC CYTOKINES. ENHANCES THE SURFACE EXPRESSION OF THE

```



```

CC INTRACELLULAR ADHESION MOLECULE-1 (ICAM-1) IN FIBROBLASTS.
CC -1 SUBUNIT: HOMODIMER.
CC -1 SUBCELLULAR LOCATION: Secreted.
CC -1 TISSUE SPECIFICITY: RESTRICTED TO ACTIVATED MEMORY T-CELLS.
CC -1 PTM: FOUND BOTH IN GLYCOSYLATED AND NONGLYCOSYLATED FORMS.
CC -1 SIMILARITY: BELONGS TO THE IL-17 FAMILY.
CC -1 DATABASE: NAME=Rad Systems' cytokine mini-reviews: IL17;
CC WWW="http://www.radsystems.com/asp/q_sitedb/asp?bodyid=211".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z58820; CAA91233.1; -.
DR EMBL: U32659; AAC50341.1; -.
DR Genew: HGNC:5981; IL17.
DR MIM: 603149; -.
KM Cytokine: Glycoprotein: T-cell; Antigen; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 135 INTERLEUKIN-17.
FT DISULEID 94 144 BY SIMILARITY.
FT DISULEID 99 146 BY SIMILARITY.
FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 155 AA; 17504 MW; 2BCAE9CB2F486D1 CRC64;

Query Match 41.7%; Score 5; DB 1; Length 155;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LNTHN 8
Db 49 LNTHN 53

RESULT 13
UBCX_PICAN STANDARD; PRT; 188 AA.
AC 060015;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ubiquitin-conjugating enzyme E2-21 kDa (EC 6.3.2.19)
DE (Ubiquitin-protein ligase) (Ubiquitin carrier protein) (Peroxin-4).
GN PEX4.
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_Taxid=4905;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 4732;
RX MEDLINE=98315053; PubMed=9649431;
RA van der Klei I.J., Halbrands R.E., Kiel J.A.K.W., Rasmussen S.W.,
RA Cregg J.M., Veenhuis M.;
RT "The ubiquitin-conjugating enzyme Pex4 of Hansenula polymorpha is
RT required for efficient functioning of the Pts1 import machinery.";
RL EMBO J. 17:3608-3618(1998).
CC -1 FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
CC PROTEINS. ESSENTIAL FOR PEROXISOME BIOGENESIS.
CC -1 CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine = AMP +
CC diphosphate + protein N-ubiquityllysine.
CC -1 PATHWAY: Ubiquitin conjugation; second step.
CC -1 MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
CC THIOLESTER FORMATION.
CC -1 SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF061604; AAC16238.1; -.
DR HSSP: P15731; LOCO.
DR InterPro: IPR000608; UBQ_conjugat.
DR Pfam: PF00179; UQ_con; 1.
DR ProDom: PD000461; UBQ_conjugat; 1.
DR SMART: SM00212; UBQC; 1.
DR PROSITE: PS00183; UBIQUITIN_CONJUGAT_1; 1.
DR PROSITE: PS0127; UBIQUITIN_CONJUGAT_2; 1.
KM Ub1 conjugation pathway: Ligase; Multigene family; Peroxisome.
FT BINDING 119 119 UBIQUITIN (BY SIMILARITY).
SQ SEQUENCE 188 AA; 21531 MW; 0BC293A5D0787C88 CRC64;

Query Match 41.7%; Score 5; DB 1; Length 188;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SPLNT 6
Db 151 SPLNT 155

RESULT 14
PGCL_PIG STANDARD; PRT; 193 AA.
AC 095250;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane associated progesterone receptor component 1.
DE GPRMC1 OR GPRMC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Vascular smooth muscle;
RX MEDLINE=97112407; PubMed=8954087;
RA Falkenstein E., Meyer C., Eisen C., Scriba P.C., Wehling M.;
RT "Full-length cDNA sequence of a progesterone membrane-binding protein
RT from porcine vascular smooth muscle cells.";
RL Biochem. Biophys. Res. Commun. 229:86-89(1996).
RN [2]
RP SEQUENCE OF 1-20, AND CHARACTERIZATION.
RC TISSUE=Liver;
RX MEDLINE=96370823; PubMed=8774719;
RA Meyer C., Schmid R., Scriba P.C., Wehling M.;
RT "Purification and partial sequencing of high-affinity progesterone-
RT binding site(s) from porcine liver membranes.";
RL Eur. J. Biochem. 239:726-731(1996).
CC -1 FUNCTION: RECEPTOR FOR PROGESTERONE.
CC -1 SUBUNIT: HETERODIMER OR HETEROTRIMER (POTENTIAL).
CC -1 SUBCELLULAR LOCATION: MICROSMAL; MEMBRANE-BOUND.
CC -1 SIMILARITY: BELONGS TO THE MAPR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X99714; CAA68050.1; -.
DR InterPro: IPR001199; Cyt_B5.
DR Pfam: PF00173; heme_1; 1.
KM Receptor; Steroid-binding; Transmembrane; Microsome.
FT INIT_MET 0 0

```

FT TRANSMEM 24 42 POTENTIAL.
SQ SEQUENCE 193 AA; 21478 MW; F7524857C95CDB23 CRC64;

Query Match 41.7%; Score 5; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPLN 5
|||||
Db 26 TSPLN 30

RESULT 15

MINC_HELPJ

ID MINC_HELPJ STANDARD; PRT; 194 AA.

AC 092M51;

DT 16+OCT-2001 (Rel. 40, Created)

DT 16+OCT-2001 (Rel. 40, Last sequence update)

DE 16+OCT-2001 (Rel. 40, Last annotation update)

GN Probable septum site-determining protein minc.

NC OR JHP0372.

OS Helicobacter pylori J99 (Campylobacter pylori J99).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

OC Helicobacter.

OX NCBI_TaxID=85963;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99120557; PubMed=9923682;

RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,

RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,

RA Tummino P.J., Caruso A., Ulla-Nickelsen M., Mills D.M., Ives C.,

RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,

RA Trust T.J.;

RT "Genomic sequence comparison of two unrelated isolates of the human

RT gastric pathogen Helicobacter pylori.";

RL Nature 387:176-180(1999).

CC -! FUNCTION: CELL DIVISION INHIBITOR THAT BLOCKS THE FORMATION OF

CC POLAR 2 RING SEPTUMS. RAPIDLY OSCILLATES BETWEEN THE POLES OF THE

CC CELL TO DESTABILIZE FTSZ FILAMENTS THAT HAVE FORMED BEFORE THEY

CC MATURE INTO POLAR 2 RINGS. PREVENT FTSZ POLYMERIZATION (BY

CC SIMILARITY)

CC -! SUBUNIT: INTERACTS WITH MIND AND FTSZ (BY SIMILARITY).

CC -! SIMILARITY: BELONGS TO THE MINC FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC of send an email to license@isb-sib.ch).

CC -----

CC EMBL: AE001472; AAD05953.1; -

CC DR Cell division; Septation; Complete proteome.

CC KW SEQUENCE 194 AA; 22302 MW; 33093F48637D0FA8 CRC64;

SQ

OY 5 NINNG 9

Db 117 NINNG 121

Query Match 41.7%; Score 5; DB 1; Length 194;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: January 3, 2003, 14:27:02
Job time : 11 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:26:33 ; Search time 16 Seconds
(without alignments)
72.101 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 12
Sequence: 1 TSPLEINHGOKL 12

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR-73:*
1: pirl:*
2: pirl:*
3: pirl:*
4: pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50.0	169	2	G83075	type 4 fimbrial bi
2	50.0	294	2	C69956	phosphate ABC tran
3	50.0	391	2	F64366	flavoprotein - Met
4	50.0	481	2	T32260	hypothetical prote
5	50.0	706	2	S42174	NPase - African s
6	50.0	706	2	QJ0210	probable nucleic a
7	50.0	1050	2	T18300	serine/threonine-s
8	50.0	1273	2	T38292	hypothetical prote
9	50.0	1690	2	T31670	DNA-directed RNA p
10	50.0	61	2	AD2424	hypothetical prote
11	50.0	63	2	D81307	hypothetical prote
12	50.0	69	2	D97809	hypothetical prote
13	50.0	76	2	T17971	hypothetical prote
14	50.0	80	2	S77069	hypothetical prote
15	50.0	100	2	B71102	hypothetical prote
16	50.0	111	2	S56800	probable membrane
17	50.0	113	2	S23653	sensorin A - Calif
18	50.0	124	2	C90205	conserved hypotet
19	50.0	124	2	H87616	cytochrome c famil
20	50.0	127	2	T44119	hypothetical prote
21	50.0	127	2	JS0155	lipid transfer pro
22	50.0	127	2	UT0590	POX18cm protein -
23	50.0	128	2	C84708	probable MYB famli
24	50.0	132	2	AD1648	hypothetical prote
25	50.0	135	2	H71051	hypothetical prote
26	50.0	137	1	SSKV	cupredoxin [valida
27	50.0	139	2	AI3059	hypothetical prote
28	50.0	139	2	G98226	hypothetical prote
29	50.0	145	2	B96705	unknown protein, 8

30	50.0	150	2	A11968	hypothetical prote
31	50.0	164	2	T16783	hypothetical prote
32	50.0	180	2	F99844	conserved hypotet
33	50.0	182	2	T10484	blue copper protei
34	50.0	182	2	G69418	probable transpos
35	50.0	184	2	A85036	hypothetical prote
36	50.0	190	2	T33459	hypothetical prote
37	50.0	194	2	H71939	hypothetical prote
38	50.0	194	2	JC5260	progestosterone membr
39	50.0	197	2	R82282	conserved hypotet
40	50.0	204	2	T51012	hypothetical prote
41	50.0	208	2	T41710	hypothetical prote
42	50.0	213	2	S72625	hypothetical prote
43	50.0	217	2	E64651	hypothetical prote
44	50.0	219	2	AD1471	ABC transporters (
45	50.0	226	2	T08335	hypothetical prote

ALIGNMENTS

RESULT 1
G83075
Type 4 fimbrial biogenesis protein FlmT PA4549 [Imported] - Pseudomonas aeruginosa (s
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83075
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; L
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: G83075
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-169 <STO>
A:Cross-references: GB:AE004869; GB:AE004091; NID:g9950793; PIDN:AA07937.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: flmT; PA4549

Query Match 50.0%; Score 6; DB 2; Length 169;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LNING 9
Db 123 LNING 128

RESULT 2
C69956
phosphate ABC transporter (permease) homolog yggI - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C:Accession: C69956
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauder, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Dancin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: C69956
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-294 <RUN>
A:Cross-references: GB:299116; GB:AL009126; NID:g2634723; PIDN:CAB14427.1; PID:g2634930
A:Experimental source: strain 168
C:Genetics:
A:Gene: y9q1
C:Superfamily: phow protein

Query Match 50.0%; Score 6; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPINT 6
|||||
Db 230 TSPINT 235

RESULT 3
F64366
flavoprotein - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 18-Sep-1998
C:Accession: F64366
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Rison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: F64366
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-391 <BUL>
A:Cross-references: GB:067502; GB:L77117; NID:g1591223; PID:g1591237; TIGR:MJ0534; PID:g
C:Genetics:
A:Map position: FOR469119-470294
A:Start codon: GTG
C:Superfamily: Methanobacterium flavoprotein A
C:Keywords: Flavoprotein

Query Match 50.0%; Score 6; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPINT 6
|||||
Db 283 TSPINT 288

RESULT 4
T32260
hypothetical protein C24A1.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 02-Sep-2000
C:Accession: T32260
R:Connell, M.
Submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid C24A1.
A:Reference number: Z21141
A:Accession: T32260
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-481 <CON>
A:Cross-references: EMBL:AF024491; PIDN:AAB70311.1; GSPDB:GN00021; CESP:C24A1.2
A:Experimental source: strain Bristol N2; clone C24A1
C:Genetics:
A:Gene: CESP:C24A1.2
A:Map position: 3
A:Introns: 55/3; 96/1; 129/2; 190/2; 209/3; 233/1; 277/1; 428/3; 445/3
C:Superfamily: ets DNA-binding domain homology

F:340-420/Domain: ets DNA-binding domain homology <ETS>

Query Match 50.0%; Score 6; DB 2; Length 481;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPINT 6
|||||
Db 21 TSPINT 26

RESULT 5
S42174
NPase - African swine fever virus
C:Species: African swine fever virus, ASFV
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C:Accession: S42174
R:Freije, J.M.P.; Lain, S.; Vinuela, E.; Lopez-otín, C.
Virus Res. 30, 63-72, 1993
A:Title: Nucleotide sequence of a nucleotide triphosphate phosphohydrolase gene from
A:Reference number: S42174; MUID:94091056; PMID:8266720
A:Accession: S42174
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-706 <PRE>
A:Cross-references: EMBL:X69952; NID:g407358; PIDN:CAA49573.1; PID:g407359

Query Match 50.0%; Score 6; DB 2; Length 706;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SPLINT 7
|||||
Db 485 SPLINT 490

RESULT 6
J02210
probable nucleic acid-dependent ATPase - African swine fever virus
N:Alternate names: helicase homolog j10L protein
C:Species: African swine fever virus, ASFV
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: J02210; S27894
R:Baylis, S.A.; Twigg, S.R.P.; Vydelingum, S.; Dixon, L.K.; Smith, G.L.
J. Gen. Virol. 74, 1969-1974, 1993
A:Title: Three African swine fever virus genes encoding proteins with homology to put
A:Reference number: J02209; MUID:93389456; PMID:8376971
A:Accession: J02210
A:Molecule type: DNA
A:Residues: 1-706 <BAV>
A:Cross-references: EMBL:X72952; NID:g414092; PIDN:CA51457.1; PID:g414093
R:Roberts, P.C.; Lu, Z.; Rock, D.L.
Submitted to the EMBL Data Library, July 1992
A:Description: Three adjacent genes of African swine fever virus with homologies to e
A:Reference number: S27892
A:Accession: S27894
A:Molecule type: DNA
A:Residues: 1-706 <ROB>
A:Cross-references: EMBL:M88275; NID:g210602; PIDN:AAA03222.1; PID:g210605
C:Genetics:
A:Gene: j10L

Query Match 50.0%; Score 6; DB 2; Length 706;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SPLINT 7
|||||
Db 485 SPLINT 490

RESULT 7
T18300

serine/threonine-specific protein kinase homolog - *Emmericella nidulans*
C:Species: *Emmericella nidulans*, *Aspergillus nidulans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18300
R:Ramos, P.; Ellis, L.
submitted to the EMBL Data Library, April 1995
A:Reference number: Z18868
A:Accession: T18300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1050 <RAM>
A:Cross-references: EMBL:U25693; NID:g818862; PID:g818863; PIDN:AAA67069.1
C:Genetics:
A:introns: 626/1; 808/3
A:Note: anka

Query Match 50.0%; Score 6; DB 2; Length 1050;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 PLNIN 8
 |||||
Db 661 PLNIN 666

RESULT 8
T38292
hypothetical protein SPAC23E2.02 - fission yeast (*Schizosaccharomyces pombe*)
C:Species: *Schizosaccharomyces pombe*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38292
R:Skellton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z21784
A:Accession: T38292
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1273 <SKR>
A:Cross-references: EMBL:Z68887; PIDN:CAA93114.1; GSPDB:GN00066; SPDB:SPAC23E2.02
A:Experimental source: strain 972h-; cosmid c23E2
C:Genetics:
A:Gene: SPDB:SPAC23E2.02
A:Map position: 1
A:introns: 8/2; 862/3

Query Match 50.0%; Score 6; DB 2; Length 1273;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 SPLNIN 7
 |||||
Db 790 SPLNIN 795

RESULT 9
T31670
DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - *Mastigamoeba invertens* (frag
C:Species: *Mastigamoeba invertens*
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C:Accession: T31670
R:Stiller, J.W.; Duffield, E.C.S.; Hall, B.D.
Proc. Natl. Acad. Sci. U.S.A. 95, 11769-11774, 1998
A:Title: Amliochondriate amoebae and the evolution of DNA-dependent RNA polymerase II.
A:Reference number: Z21052; MUID:98426228; PMID:9751740
A:Accession: T31670
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1690 <STL>
A:Cross-references: EMBL:AF083338; NID:g3511286; PID:g3511287; PIDN:AAC62246.1
A:Experimental source: ATCC50338
C:Genetics:
A:Gene: RPBI
A:Note: Intron positions not resolved (incomplete sequence)

C:Superfamily: human DNA-directed RNA polymerase II largest chain
C:Keywords: DNA binding; nucleotidyltransferase; transcription

Query Match 50.0%; Score 6; DB 2; Length 1690;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 SPLNIN 7
 |||||
Db 294 SPLNIN 299

RESULT 10
AD2424
hypothetical protein asl4948 [imported] - *Nostoc* sp. (strain PCC 7120)
C:Species: *Nostoc* sp.
A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AD2424
R:Kaneke, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2424
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-61 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB76647.1; PID:g17134086; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: asl4948

Query Match 41.7%; Score 5; DB 2; Length 61;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TSPIN 5
 |||||
Db 33 TSPIN 37

RESULT 11
D81307
hypothetical protein Cj1047c [imported] - *Campylobacter jejuni* (strain NCTC 11168)
C:Species: *Campylobacter jejuni*
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: D81307
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chli
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Bar
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals
A:Reference number: AB1250; MUID:20150912; PMID:10688204
A:Accession: D81307
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-63 <PAR>
A:Cross-references: GB:AL139077; GB:AL111168; NID:g6968444; PIDN:CAB73303.1; PID:g696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj1047c

Query Match 41.7%; Score 5; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 NGORKL 12
 |||||
Db 4 NGORKL 8

RESULT 12
D97809

hypothetical protein RC0876 [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: D97809
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fourmter, P.E.; Barbe, V.; Samson, D.; R
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: D97809
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-69 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03414.1; PID:g15619981; GSPDB:GN00173
C:Genetics:
A:Gene: RC0876

Query Match 41.7%; Score 5; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HNGOK 11
| | | | |
Db 23 HNGOK 27

RESULT 13
T17971
hypothetical protein a469L - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17971
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T17971
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-76 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC36836.1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:
A:Note: a469L

Query Match 41.7%; Score 5; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPLN 5
| | | | |
Db 56 TSPLN 60

RESULT 14
S77069
hypothetical protein sll1263 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S77069
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimo, S.; Takeduchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S77069
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-80 <KAN>
A:Cross-references: EMBL:D64005; GB:AB001339; NID:g1001779; PIDN:BAI0761.1; PID:dt01141
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 41.7%; Score 5; DB 2; Length 80;

Best Local Similarity 100.0%; Pred. No. 57;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPLN 5
| | | | |
Db 60 TSPLN 64

RESULT 15
B71102
hypothetical protein PH1077 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: B71102
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
DNA Res 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: B71102
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-100 <KAN>
A:Cross-references: GB:AP000004; NID:g3236131; PIDN:BAJ0176.1; PID:g3257493
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by Genba
C:Genetics:
A:Gene: PH1077
C:superfamily: Pyrococcus horikoshii hypothetical protein PH1077

Query Match 41.7%; Score 5; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPLN 5
| | | | |
Db 83 TSPLN 87

Search completed: January 3, 2003, 14:28:42
Job time : 17 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:24:57 ; Search time 11 Seconds
(without alignments)
20.674 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 64
Sequence: 1 TSPLNHNGOKL 12

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubppaa/PC7_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubppaa/US06_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubppaa/PC7US_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
- 9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
- 11: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64	100.0	12	US-09-899-376-1	Sequence 1, Appl1
2	64	100.0	19	US-09-899-376-3	Sequence 3, Appl1
3	64	100.0	19	US-09-899-376-4	Sequence 4, Appl1
4	42	65.6	1325	US-09-741-669-304	Sequence 304, App
5	40	62.5	61	US-09-864-761-40714	Sequence 40714, A
6	38	59.4	247	US-10-042-417-36	Sequence 36, Appl1
7	37	57.8	240	US-09-815-242-5066	Sequence 5066, Ap
8	36	56.2	194	US-09-815-242-11500	Sequence 11500, A
9	36	56.2	196	US-09-815-242-11660	Sequence 11660, A
10	36	56.2	252	US-09-815-242-10296	Sequence 10296, A
11	36	56.2	259	US-09-815-242-13738	Sequence 13738, A
12	35	54.7	86	US-09-864-761-39192	Sequence 39192, A
13	35	54.7	160	US-09-864-761-47255	Sequence 47255, A
14	35	54.7	1088	US-09-920-804-2	Sequence 2, Appl1
15	35	54.7	2000	US-10-010-901-29	Sequence 29, Appl1
16	35	54.7	2150	US-10-135-322-17	Sequence 17, Appl1
17	34	53.1	287	US-09-800-729-114	Sequence 114, App
18	34	53.1	287	US-09-800-729-153	Sequence 153, App
19	34	53.1	332	US-09-925-297-786	Sequence 786, App

20	34	53.1	478	10	US-09-417-791-5	Sequence 5, Appl1
21	34	53.1	558	10	US-09-864-761-47724	Sequence 47724, A
22	34	53.1	613	10	US-09-800-729-82	Sequence 82, Appl1
23	34	53.1	613	10	US-09-800-729-98	Sequence 98, Appl1
24	33	51.6	333	10	US-09-864-761-34602	Sequence 34602, A
25	33	51.6	209	10	US-09-764-870-323	Sequence 323, App
26	33	51.6	209	10	US-09-764-853-546	Sequence 546, App
27	33	51.6	278	10	US-09-903-456-18	Sequence 18, Appl1
28	33	51.6	337	9	US-10-063-547-74	Sequence 74, Appl1
29	33	51.6	337	12	US-10-006-867-74	Sequence 74, Appl1
30	33	51.6	337	12	US-10-052-586-268	Sequence 268, App
31	33	51.6	342	10	US-09-864-761-36708	Sequence 36708, A
32	33	51.6	345	10	US-09-815-242-11499	Sequence 11499, A
33	33	51.6	505	10	US-09-815-242-5857	Sequence 5857, Ap
34	33	51.6	505	10	US-09-815-242-12868	Sequence 12868, A
35	33	51.6	870	10	US-09-815-242-5493	Sequence 5493, Ap
36	33	51.6	870	10	US-09-815-242-12637	Sequence 12637, A
37	33	51.6	2139	10	US-09-727-384-6	Sequence 6, Appl1
38	33	51.6	2697	10	US-09-961-527A-5	Sequence 5, Appl1
39	33	51.6	3571	9	US-10-150-821-2	Sequence 2, Appl1
40	33	51.6	3571	10	US-09-911-842-2	Sequence 2, Appl1
41	32	50.8	677	10	US-09-815-242-11921	Sequence 11921, A
42	32	50.0	10	10	US-09-984-056-11	Sequence 11, Appl1
43	32	50.0	10	10	US-09-984-057-11	Sequence 11, Appl1
44	32	50.0	30	10	US-09-989-789-9	Sequence 9, Appl1
45	32	50.0	60	10	US-09-764-877-1853	Sequence 1853, Ap

ALIGNMENTS

RESULT 1
US-09-899-376-1
Sequence 1, Application US/09899376
Patent No. US20020102265A1
GENERAL INFORMATION:
APPLICANT: HONG, FRANK D.
TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT INFILTRATES TUMOR TISSUE FOR TARGETED DRUG DELIVERY
FILE REFERENCE: UTSC:64505
CURRENT FILING DATE: US/09/899, 376
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 12
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-899-376-1

Query Match 100.0%; Score 64; DB 10; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.3e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPLNHNGOKL 12
|||||
Db 1 TSPLNHNGOKL 12

RESULT 2
US-09-899-376-3
Sequence 3, Application US/09899376
Patent No. US20020102265A1
GENERAL INFORMATION:
APPLICANT: HONG, FRANK D.
TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT INFILTRATES TUMOR TISSUE FOR TARGETED DRUG DELIVERY
FILE REFERENCE: UTSC:64505

```

: CURRENT APPLICATION NUMBER: US/09/899,376
: CURRENT FILING DATE: 2001-07-02
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3
: LENGTH: 19
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Synthetic
: OTHER INFORMATION: Peptide
US-09-899-376-3

Query Match          100.0%; Score 64; DB 10; Length 19;
Best Local Similarity 100.0%; Pred. No. 5, 6e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TSPLNHNGQKL 12
Db      4 TSPLNHNGQKL 15

RESULT 3
US-09-899-376-4
: Sequence 4, Application US/09899376
: Patent No. US20020102265A1
: GENERAL INFORMATION:
: APPLICANT: HONG, FRANK D.
: APPLICANT: CLAYMAN, GARY.
: TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT
: FILE REFERENCE: US:64505
: CURRENT APPLICATION NUMBER: US/09/899,376
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 4
: LENGTH: 19
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Synthetic
: OTHER INFORMATION: Peptide
US-09-899-376-4

Query Match          100.0%; Score 64; DB 10; Length 19;
Best Local Similarity 100.0%; Pred. No. 5, 6e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TSPLNHNGQKL 12
Db      8 TSPLNHNGQKL 19

RESULT 4
US-09-741-669-304
: Sequence 304, Application US/09741669
: Patent No. US20020022718A1
: GENERAL INFORMATION:
: APPLICANT: Forsyth, R. Allyn
: APPLICANT: Ohlson, Karl L.
: APPLICANT: Zyskind, Judith W.
: TITLE OF INVENTION: Genes identified as required for
: FILE REFERENCE: ELTRA.009A
: CURRENT APPLICATION NUMBER: US/09/741,669
: CURRENT FILING DATE: 2000-12-19
: PRIOR APPLICATION NUMBER: US 60/173005
: PRIOR FILING DATE: 1999-12-23
: NUMBER OF SEQ ID NOS: 481
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 304
: LENGTH: 1325
```

```

: TYPE: PRT
: ORGANISM: Escherichia coli
US-09-741-669-304

Query Match          65.6%; Score 42; DB 10; Length 1325;
Best Local Similarity 88.9%; Pred. No. 8.4;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 LNHNGQKL 12
Db      991 LNHNGQKL 999

RESULT 5
US-09-864-761-40714
: Sequence 40714, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
: FILE REFERENCE: Aeomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 40714
: LENGTH: 61
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO U82670.2
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.7
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.5
```



```
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
; OTHER INFORMATION: SWISSPROT HIT: P94598, EVALUO 1.50e+00
US-09-864-761-40714
```

```
Query Match
Best Local Similarity 72.5%; Score 40; DB 10; Length 61;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 TSPLNHNGOKL 11
    ||| :| ||||
Db 35 TSPKHNGOKL 45
```

```
RESULT 6
US-10-042-417-36
; Sequence 36, Application US/10042417
; Patent No. US20020123082A1
; GENERAL INFORMATION:
; APPLICANT: Pagano, M.
; TITLE OF INVENTION: METHODS TO IDENTIFY COMPOUNDS USEFUL FOR THE TREATMENT OF
; FILE REFERENCE: 5914-090-999
; CURRENT APPLICATION NUMBER: US/10/042,417
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: 60/260,179
; PRIOR FILING DATE: 2001-01-5
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-417-36
```

```
Query Match
Best Local Similarity 100.0%; Score 38; DB 12; Length 247;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 6 IHNGOKL 12
    |||||
Db 146 IHNGOKL 152
```

```
RESULT 7
US-09-815-242-5066
; Sequence 5066, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
```

```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5066
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5066
```

```
Query Match
Best Local Similarity 57.8%; Score 37; DB 10; Length 240;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY 2 SPLNHNHNGOKL 12
    ||| ||| |||
Db 75 SPLNHNHNGOKL 85
```

```
RESULT 8
US-09-815-242-11500
; Sequence 11500, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11500
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11500
```

```
Query Match
Best Local Similarity 56.2%; Score 36; DB 10; Length 194;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 5 NIHNHNGOKL 12
    |||||
Db 117 NIHNHNGOKL 124
```

RESULT 9

```
US-09-815-242-11660
; Sequence 11660, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11660
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-815-242-11660

Query Match          56.2%; Score 36; DB 10; Length 196;
Best Local Similarity 77.8%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SPLNHNGQ 10
Db 29 SPLHHLGQ 37

RESULT 10
US-09-815-242-10296
; Sequence 10296, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13738
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13738

Query Match          56.2%; Score 36; DB 10; Length 259;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10296
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10296

Query Match          56.2%; Score 36; DB 10; Length 252;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SPLNHNGQ 10
Db 85 SPLHHLGQ 93

RESULT 11
US-09-815-242-13738
; Sequence 13738, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13738
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13738

Query Match          56.2%; Score 36; DB 10; Length 259;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 SPLNHNGQ 10
Db 92 SPLHHLGQ 100
```

```
RESULT 12
US-09-864-761-39192
; Sequence 39192, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39192
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC003684.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.3
; OTHER INFORMATION: SWISSPROT HIT: O70311, EVALUATE 3.90e+00
US-09-864-761-39192
```

Query Match 54.7%; Score 35; DB 10; Length 86;
Best Local Similarity 62.5%; Pred. No. 7.5;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 NINHGOKL 12
|||||
DB 38 NVHNGSKV 45

```
RESULT 13
US-09-864-761-47255
; Sequence 47255, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aemica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47255
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002310.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.52
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.53
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: SWISSPROT HIT: P52742, EVALUATE 1.00e-55
; OTHER INFORMATION: EST_HUMAN HIT: BF590267.1, EVALUATE 4.00e-99
US-09-864-761-47255
```

Query Match 54.7%; Score 35; DB 10; Length 160;
Best Local Similarity 54.5%; Pred. No. 15;

Matches	6, Conservative	2, Mismatches	3, Indels	0, Gaps
QY	1 TSPLINHGQK 11			
	:			
Db	70 TSHLRVHTGEK 80			

```

RESULT 14
US-09-920-804-2
: Sequence 2, Application US/09920804
: Patent No. US20020064845A1
: GENERAL INFORMATION:
: APPLICANT: Seino, Susumu, JCR Pharmaceuticals Co., Ltd.
: TITLE OF INVENTION: Sodium Ion-Driven Chloride/Bi-Carbonate Exchanger
: FILE REFERENCE: GP44
: CURRENT APPLICATION NUMBER: US/09/920,804
: CURRENT FILING DATE: 2001-08-03
: NUMBER OF SEQ ID NOS: 8
: SEQ ID NO 2
: LENGTH: 1088
: TYPE: PRT
: ORGANISM: Mus musculus
: US-09-920-804-2

```

Query/Match	54.7%	Score 35:	DB 10:	length 1088;
Best local similarity	50.0%	Pred. No.	1.3e+02;	
Matches	6:	Conservative	3:	Mismatches 3:
				Indels 0:
				Gaps 0:

```
QY      1 TSP LNHGQKL 12
        | | | | |
Db      620 TYP INMHNDLEL 631
```

```

RESULT 15
US-10-010-901-29
: Sequence 29, Application US/10010901
: Patent No. US20020096201A1
: GENERAL INFORMATION:
: APPLICANT: McFadden, Grant
: TITLE OF INVENTION: NOVEL MYXOMA GENES FOR IMMUNE MODULATION
: FILE REFERENCE: 50082/009002
: CURRENT APPLICATION NUMBER: US/10/010,901
: CURRENT FILING DATE: 2001-12-06
: PRIOR APPLICATION NUMBER: US/09/615,041
: PRIOR FILING DATE: 2000-07-12
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 29
: LENGTH: 2000
: TYPE: PR1
: ORGANISM: Myxoma Virus
: US-10-010-901-29

```

Query/Match	54.7%	Score 35;	DB 12;	Length 2000;
Best local similarity	54.5%	Pred. NO.	2.7e+02;	
Matches	6;	Conservative	2;	Mismatches 3;
				Indels 0;
				Gaps 0;

Qy	1	TSP	LNI	HNG	QK	11
Db	230	TLP	VS	IHQ	GSK	240

```
Search completed: January 3, 2003, 14:26:44
Job time : 11 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:23:57 ; Search time 29 Seconds

(without alignments)
85.261 Million cell updates/sec

Title: US-09-899-376-1

Perfect score: 64

Sequence: 1 TSPININGOKL 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_rv1rus:*
17: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	67.2	321	4	Q96ER2
2	42	65.6	1343	16	Q8XAY4
3	42	65.6	2531	5	016004
4	41	64.1	427	10	Q9SZV7
5	40	62.5	331	3	014097
6	40	62.5	334	5	09GUZ8
7	39	60.9	243	16	Q8ZHG4
8	39	60.9	392	4	Q9H807
9	39	60.9	563	16	Q9X8S9
10	39	60.9	738	4	Q9U1A1
11	39	60.9	855	17	Q8TF99
12	39	60.9	914	12	Q9IF30
13	39	60.9	938	12	Q9W8S1
14	39	60.9	939	12	Q84178
15	39	60.9	970	5	Q96115
16	39	60.9	2176	5	Q9VGP1

17	38	59.4	147	11	Q9CUI6	Q9CUI6 mus musculi
18	38	59.4	150	16	Q9CUI1	Q9CUI1 lactococcus
19	38	59.4	173	4	Q9UKC5	Q9UKC5 homo sapien
20	38	59.4	188	2	Q93W11	Q93W11 photorhabd
21	38	59.4	277	2	Q8VTA6	Q8VTA6 synechococ
22	38	59.4	324	5	Q22161	Q22161 caenorhabd1
23	38	59.4	367	11	Q91V19	Q91V19 mus musculi
24	38	59.4	376	4	Q96CK9	Q96CK9 homo sapien
25	38	59.4	406	16	Q8ZMW4	Q8ZMW4 salmonella
26	38	59.4	406	16	Q8Z416	Q8Z416 salmonella
27	38	59.4	415	11	Q9JTE4	Q9JTE4 mus musculi
28	38	59.4	444	10	P93716	P93716 petunia hyb
29	38	59.4	471	4	Q9H0V2	Q9H0V2 homo sapien
30	38	59.4	480	11	Q9DC63	Q9DC63 mus musculi
31	38	59.4	480	11	Q9D2J5	Q9D2J5 mus musculi
32	38	59.4	641	5	Q18586	Q18586 caenorhabd1
33	38	59.4	693	10	Q9SKH6	Q9SKH6 arabidopsis
34	37	57.8	111	16	Q8XAE1	Q8XAE1 escherichia
35	37	57.8	112	4	Q9H4Y3	Q9H4Y3 homo sapien
36	37	57.8	177	5	Q95S82	Q95S82 drosophila
37	37	57.8	240	16	Q91694	Q91694 pseudomonas
38	37	57.8	248	2	Q87957	Q87957 salmonella
39	37	57.8	248	12	Q9J8C4	Q9J8C4 spodiopetra
40	37	57.8	299	2	Q9F7K7	Q9F7K7 salmonella
41	37	57.8	299	2	Q9F7K4	Q9F7K4 salmonella
42	37	57.8	299	2	Q9F7K0	Q9F7K0 salmonella
43	37	57.8	299	2	Q9F7J7	Q9F7J7 salmonella
44	37	57.8	299	2	Q9F7I9	Q9F7I9 salmonella
45	37	57.8	299	2	Q9F7I4	Q9F7I4 salmonella

ALIGNMENTS

RESULT 1

Q96ER2 ID Q96ER2 PRELIMINARY; PRT; 321 AA.
AC Q96ER2;
DT 01-DEC-2001 (TREMBLrel. 19, created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Unknown (protein for MGC:21259).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS, AND EMBRYONAL CARCINOMA;
RA Strauberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC012012; AAH12012.1; -.
DR InterPro; IPR001909; KRA8.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 7.
DR ProDom; PD000003; Znf_C2H2; 6.
DR PROSITE; PS50805; KRA8; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_6.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 321 AA; 36369 MW; 62BE342C8B7067D9 CRC64;

Query Match 67.2%; Score 43; DB 4; Length 321;

Best local Similarity 80.0%; Pred. No. 5.3;

Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 3 PLININGOKL 12
| | | | |
Db 310 PKRININGOKL 319

RESULT 2

```

08XAY4
ID 08XAY4 PRELIMINARY; PRT; 1343 AA.
AC 08XAY4;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Of, hypothetical protein.
GN YDEK OR Z2195 OR ECS2117.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.U., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boulton A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lam A., Dimalanta E.T., Potamoukis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blatter F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shida T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL: AE005354; AAG56256.1; -
DR EMBL: AP002557; BAB35540.1; -
DR Complete proteome.
SQ SEQUENCE 1343 AA; 138568 MW; 4C2456739907B63D CRC64;

Query Match 65.6%; Score 42; DB 16; Length 1343;
Best Local Similarity 88.9%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNIHNGQKL 12
DB 991 LNIHNGQKL 999

RESULT 3
016004 PRELIMINARY; PRT; 2531 AA.
ID 016004;
AC 016004;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Notch homolog.
OS Lytechinus variegatus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinozoa; Echinacea; Temnopleurozoa; Toxopneustidae;
OC Lytechinus.
OX NCBI_TaxID=7654;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97454256; PubMed=9310331;
RA Sherwood D.R., McClay D.R.;
RT Identification and localization of a sea urchin Notch homologue:
RT insights into vegetal plate regionalization and Notch receptor
RT regulation.";
RL Development 124:3363-3374(1997).
DR EMBL: AF000634; AAB82088.1; -
DR HSSP: P01132; IEGF.
DR InterPro: IPR002110; ANK.

```

```

DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_11.
DR InterPro: IPR000800; Notch.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF; 35.
DR Pfam: PF00066; notch; 3.
DR PRINTS: PR01415; ANKRIN.
DR PRINTS: PR00010; EGFBL00D.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 5.
DR SMART: SM00179; EGF_CA; 23.
DR SMART: SM00001; EGF_Like; 11.
DR SMART: SM00004; NL; 3.
DR PROSITE: PS50088; ANK_REPEAT; 10.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 2.
DR PROSITE: PS00010; ASX_HYDROXYL; 21.
DR PROSITE: PS00022; EGF_1; UNKNOWN_33.
DR PROSITE: PS01186; EGF_2; 25.
DR PROSITE: PS01187; EGF_CA; 20.
DR ANK repeat: Calcium-binding; EGF-like domain; Glycoprotein;
DR Hydroxylation; Repeat.
SQ SEQUENCE 2531 AA; 273982 MW; 5BF42BEC627CA303 CRC64;

Query Match 65.6%; Score 42; DB 5; Length 2531;
Best Local Similarity 60.0%; Pred. No. 67;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPINHNGQ 10
DB 2097 TSPMDMHNGE 2106

RESULT 4
09S2V7 PRELIMINARY; PRT; 427 AA.
ID 09S2V7;
AC 09S2V7;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 47.4 kDa protein.
GN F6G3.70 OR ATG30040.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL078464; CAB43839.1; -
DR EMBL: AL161576; CAB80997.1; -
DR InterPro: IPR001461; AsproteaseA1.
DR Pfam: PF00026; asp; 1.
DR PRINTS: PR00792; PEPSTIN.
DR Hypothetical protein.
SQ SEQUENCE 427 AA; 47357 MW; F6FA2BDD938B3225 CRC64;

```

Query Match 64.1%; Score 41; DB 10; Length 427;
 Best Local Similarity 77.8%; Pred. No. 17;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TSPINHNG 9
 1:1111111111
 DB 263 TTPLEIHNG 271

RESULT 5

ID 014097 PRELIMINARY; PRT; 331 AA.
 AC 014097;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical 37.7 kDa proline-rich protein C2P5.14C in chromosome I.
 GN SPAC2P5.14C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Oliver K., Harris D., Barrell B.G., Rajandream M.A., Mood V.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z99145; CAB16268.1;
 DR InterPro: IPR001202; WW_hsp5_wmp.
 DR Pfam: PF00397; WW; 1.
 DR Hypothetical protein.
 KW Hypothetical protein.
 SQ SEQUENCE 331 AA; 37669 MW; EA888EACCA4CB7B1 CRC64;

Query Match 62.5%; Score 40; DB 3; Length 331;
 Best Local Similarity 72.7%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 SPLINHNG 12
 1111111111
 DB 271 SSLSHNGSL 281

RESULT 6

ID 09GU28 PRELIMINARY; PRT; 334 AA.
 AC 09GU28;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Brachyury protein.
 GN ODT.
 OS Oikopleura dioica.
 OC Eukaryota; Metazoa; Chordata; Trochordata; Appendicularia;
 OC Oikopleuridae; Oikopleura.
 OX NCBI_TaxID=34765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bassham S., Postlethwait J.H.;
 RT "Brachyury (Y) expression in embryos of a larvacean urochordate,
 RT Oikopleura dioica, and the ancestral role of brachyury.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF204208; AAC22592.1;
 DR HSSP: P24781; 1XBR.
 DR InterPro: IPR001699; TF_T-box.
 DR Pfam: PF00907; T-box; 1.
 DR PRINTS: PR00937; TBOX.
 DR SMART: SM00425; TBOX; 1.
 DR PROSITE: PS01283; TBOX_1; 1.
 DR PROSITE: PS50252; TBOX_3; 1.
 SQ SEQUENCE 334 AA; 37574 MW; 02AEX233B5800F50 CRC64;

Query Match 62.5%; Score 40; DB 5; Length 334;
 Best Local Similarity 58.3%; Pred. No. 21;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 TSPINHNG 12
 1:1111111111
 DB 125 TNKLHNGQOI 136

RESULT 7

ID 08ZHG4 PRELIMINARY; PRT; 243 AA.
 AC 08ZHG4;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein YPO0934.
 GN YPO0934.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Tlhalil R.W., Holden M.T.G.,
 RA Bentley S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Slimmons M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague."
 RL Nature 413:523-527(2001).
 DR EMBL: AJ414145; CAC89777.1;
 DR InterPro: IPR004382; Cons_hypoth46.
 DR TIGRFAMs: TIGR00046; Cons_hypoth46; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 243 AA; 26920 MW; BFF54E709F9A37CC CRC64;

Query Match 60.9%; Score 39; DB 16; Length 243;
 Best Local Similarity 77.8%; Pred. No. 23;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 SPLINHNG 10
 1111111111
 DB 76 SPLNHLHGQ 84

RESULT 8

ID 09H807 PRELIMINARY; PRT; 392 AA.
 AC 09H807;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE CDNA FLJ14011 fls, clone Y79AA1002472, weakly similar to zinc finger
 DE protein 91.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shitatori A., Sudo H.,
 RA Wagaatsuma M., Hosofiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nakahari K., Masuno Y.,
 RA Niomiyu K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project.";

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL: AK024073; BAB14816.1; -
 DR HSSP: P08046; 1A1H.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; Zf-C2H2; 13.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR ProDom: PD000003; Znf_C2H2; 7.
 DR SMART: SM00355; Znf_C2H2; 12.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 10.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 12.
 DR DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
 KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
 SQ SEQUENCE 392 AA; 45453 MW; 33984A6D50082D2D CRC64;

 Query Match 60.9%; Score 39; DB 4; Length 392;
 Best Local Similarity 77.8%; Pred. No. 37;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 4 LNHHNGKL 12
 Db 159 LRHHNGKL 167
 |||||:

 RESULT 9
 Q9X8S9 PRELIMINARY; PRT; 563 AA.
 AC Q9X8S9;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein SC03893.
 GN SC03893 OR SCH24.15C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 NC NCB1_TaxID=1902;
 RN 11;
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145.
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neil S.,
 RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL: AL049826; CAB42719.1; -
 KW Hypothetical protein
 SQ SEQUENCE 563 AA; 60226 MW; 7FB4DD9B5E4EB72 CRC64;

 Query Match 60.9%; Score 39; DB 16; Length 563;
 Best Local Similarity 58.3%; Pred. No. 54;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

 QY 1 TSPNHHNGKL 12
 Db 57 TSPPELHSHGRL 68
 |||||:

 RESULT 10
 Q9ULAI PRELIMINARY; PRT; 738 AA.
 AC Q9ULAI;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE R31155_1.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NC NCB1_TaxID=9606;
 RN 11;
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
 RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
 RA Pan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
 RA Dangnanan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
 RA Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
 RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
 RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
 RA Olsen A.S., Carrano A.V.;
 RT "Sequence analysis of a 2.3 Mb region containing a zinc finger (ZNF)
 cluster in 19q13.4.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL: AC013256; AAF06067.1; -
 DR HSSP: P08046; 1A1J.
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF00096; Zf-C2H2; 16.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR ProDom: PD000003; Znf_C2H2; 7.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; Znf_C2H2; 15.
 DR PROSITE: PS00805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 14.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 15.
 KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
 SQ SEQUENCE 738 AA; 84252 MW; 9E158059D5C18527 CRC64;

 Query Match 60.9%; Score 39; DB 4; Length 738;
 Best Local Similarity 77.8%; Pred. No. 71;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 4 LNHHNGKL 12
 Db 505 LRHHNGKL 513
 |||||:

 RESULT 11
 Q8TT99 PRELIMINARY; PRT; 855 AA.
 AC Q8TT99;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Peptidase family protein U32.
 GN MA0538.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 NC NCB1_TaxID=2214;
 RN 11;
 RP SEQUENCE FROM N.A.
 RC STRAIN-G2A / ANCC 35395 / DSM 2834;
 RX MEDLINE-21929760; PubMed-11932238;
 RA Galagan J.E., Nisbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Althor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thumann N., Dearellano R., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grhame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Kizycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 and physiological diversity.";
 RL Genome Res. 12:532-542(2002).

DR EMBL: AE010714; AAM03982.1; -.
KM Complete proteome.
SQ SEQUENCE 855 AA; 94070 MW; 19CF36AB92AC075 CRC64;

Query Match 60.9%; Score 39; DB 17; Length 855;
Best Local Similarity 87.5%; Pred. No. 83;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 SPLINING 9
DB 676 SPTINING 683

RESULT 12

O9IF30 PRELIMINARY; PRT; 914 AA.
AC O9IF30:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hexon protein (Fragment).
OS Bovine adenovirus type 10 (Mastadenovirus hos10).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=39788;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=78-5371;
RA Lehmkuhl H.D., Hobbs L.A.;
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF282774; AAF82136.1; -.
DR HSSP: P03277; IDHX.
DR InterPro: IPR000736; Adeno_hexon.
DR Pfam: PF01065; Adeno_hexon; 1.
DR Pfam: PF03678; Adeno_hexon.C; 1.
DR ProDom: PD002815; Adeno_hexon; 1.
FT NON-TER 914
SQ SEQUENCE 914 AA; 103905 MW; 5508E06997739CD CRC64;

Query Match 60.9%; Score 39; DB 12; Length 914;
Best Local Similarity 77.8%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 SPLINING 10
DB 217 SPTINING 225

RESULT 13

O9WBS1 PRELIMINARY; PRT; 938 AA.
ID O9WBS1:
AC O9WBS1:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE HEXON.
OS Porcine adenovirus type 3 (PAV-3).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=35265;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IAF;
RA Laroque D., Malenfant F., Massie B., Dea S.;
RT "porcine adenovirus serotype 3, complete genome."
RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA "porcine adenovirus serotype 3, complete genome."
RT Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
RL EMBL: AJ237815; CAB41030.1; -.
DR EMBL: AB026117; BAA76968.1; -.
DR HSSP: P03277; IDHX.
DR InterPro: IPR000736; Adeno_hexon.

DR Pfam: PF01065; Adeno_hexon; 1.
DR Pfam: PF03678; Adeno_hexon.C; 1.
DR ProDom: PD002815; Adeno_hexon; 1.
SQ SEQUENCE 938 AA; 105964 MW; 8115164CD481DDBA CRC64;

Query Match 60.9%; Score 39; DB 12; Length 938;
Best Local Similarity 77.8%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 SPLINING 10
DB 223 SPTINING 231

RESULT 14

O84178 PRELIMINARY; PRT; 939 AA.
ID O84178:
AC O84178:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hexon.
OS Porcine adenovirus type 3 (PAV-3).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=35265;
RN [1]
RP SEQUENCE FROM N.A.
RA McCoy R.J., Johnson M.A., Shepard M.;
RL Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL: U34592; AAB02183.1; -.
DR HSSP: P03277; IDHX.
DR InterPro: IPR000736; Adeno_hexon.
DR Pfam: PF01065; Adeno_hexon; 1.
DR Pfam: PF03678; Adeno_hexon.C; 1.
DR ProDom: PD002815; Adeno_hexon; 1.
SQ SEQUENCE 939 AA; 105897 MW; 10DB0854972DC899 CRC64;

Query Match 60.9%; Score 39; DB 12; Length 939;
Best Local Similarity 77.8%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 SPLINING 10
DB 223 SPTINING 231

RESULT 15

O96115 PRELIMINARY; PRT; 970 AA.
ID O96115:
AC O96115:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE GH21817p.
GN CG5270.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Anthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
RL EMBL: AY051570; AAK92994.1; -.
DR FlyBase: FBgn0037897; CG5270.
DR InterPro: IPR000306; Znf_FYVE.
DR Pfam: PF01363; FYVE; 1.

SQ SEQUENCE 970 AA: 108912 MW: 7FACDDDE9852607 CRC64:

Query Match 60.9%; Score 39; DB 5; Length 970;

Best Local Similarity 66.7%; Pred. No. 94;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 TSPLNINHGOKL 12

DB 374 TGPLNNHNLRL 385

Search completed: January 3, 2003, 14:26:07
Job time : 31 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:22:12 ; Search time 11 Seconds

(without alignments)
45.247 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 64
Sequence: 1 TSPLINHGOKL 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	65.6	300	RFBD_SHIFL	P37778 Shigella fl
2	42	65.6	1302	MDR5_DROME	O00748 drosophila
3	42	65.6	1325	YDER_ECOLI	P32051 escherichia
4	39	60.9	939	HEX_ADEP3	O9Ytr8 porcine ade
5	38	59.4	223	GLI4_HUMAN	P10075 homo sapien
6	38	59.4	221	Z177_HUMAN	O13360 homo sapien
7	38	59.4	471	FBX3_HUMAN	O9UK99 homo sapien
8	38	59.4	683	PLOI_SCHPO	P50528 schizosacch
9	37	57.8	111	SOHA_ECOLI	P15373 escherichia
10	37	57.8	299	RFBD_SALTY	P26392 salmonella
11	37	57.8	476	GAG_BIV06	P19558 bovine immu
12	37	57.8	476	GAG_BIV27	P19559 bovine immu
13	36	56.2	194	MINC_HELPJ	O9Zm11 helicobacte
14	36	56.2	195	MINC_HELPJ	O25653 helicobacte
15	36	56.2	243	YGJY_ECOLI	P37912 escherichia
16	36	56.2	253	Y410_BUCAI	P57488 buchnera ap
17	36	56.2	307	DNAI_SYNY3	P73097 synecocyst
18	36	56.2	430	SUN_COXBU	P45679 treponema p
19	36	56.2	605	Y596_TREPA	O83607 treponema p
20	36	56.2	1005	BGAL_ACTPL	P70753 actinobacti
21	35	54.7	288	DCMA_METSP	P21161 methylobact
22	35	54.7	455	HNA4_XENLA	O91766 xenopus lae
23	35	54.7	491	G6PD_BUCAI	P57405 buchnera ap
24	35	54.7	551	SYE_ARCFU	O29979 archaeoglob
25	35	54.7	928	YDGI_SCHPO	O10491 chloroflexu
26	35	54.7	942	DPOL_CHLAU	O08307 chloroflexu
27	34	53.1	93	YGGJ_ERWCH	P37995 erwincia chr
28	34	53.1	205	LEXA_PRORE	O07267 proviendia
29	34	53.1	223	GTXI_TOBAC	O03662 nicotiana t
30	34	53.1	287	YTFP_YEAST	P40186 saccharomyc
31	34	53.1	287	KDSA_CAUCR	O9A8C5 caulobacter
32	34	53.1	299	RBD1_ECOLI	P37760 escherichia
33	34	53.1	301	RBD2_ECOLI	O46769 escherichia

34	34	53.1	577	1	G6PC_SOLITU	O43839 solanum tub
35	34	53.1	672	1	OSM3_CABEL	P46873 caenorhabdi
36	34	53.1	721	1	BBS2_MOUSE	O9Cwf6 mus musculu
37	34	53.1	721	1	BBS2_RAT	O99m19 ratius norv
38	34	53.1	783	1	K6PF_ASPNG	P78985 aspergillus
39	34	53.1	948	1	HP1I_DEIRA	P56867 deinococcus
40	34	53.1	959	1	VILI_DICDI	P36418 dictyosteli
41	34	53.1	1033	1	CR2_HUMAN	P20023 homo sapien
42	34	53.1	1064	1	KIN1_YEAST	P13185 saccharomyc
43	34	53.1	3415	1	POLG_FOMVL	O04538 t genome po
44	33	51.6	168	1	ZG42_XENLA	P18720 xenopus lae
45	33	51.6	196	1	Y546_NETMA	O9Jw50 neisseria m

ALIGNMENTS

RESULT 1	RFBD_SHIFL	STANDARD	PRT	300 AA.
AC	P37778			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	01-OCT-1994 (Rel. 30, Last annotation update)			
DE	dtDP-4-dehydroxymannose reductase (EC 1.1.1.133) (dtDP-4-keto-L-rhamnose reductase) (dtDP-6-deoxy-L-mannose dehydrogenase) (dtDP-L-rhamnose synthetase).			
GN	RFBD.			
OS	Shigella flexneri.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OX	Shigella.			
OX	NCBI_TaxID=633;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Serotype 2A;			
RX	MEDLINE=94224146; PubMed=8170390;			
RA	Macpherson D.F., Manning P.A., Morona R.;			
RT	"Characterization of the dtDP-rhamnose biosynthetic genes encoded in the rfb locus of Shigella flexneri.";			
RL	Mol. Microbiol. 11:281-292(1994).			
CC	-1- CATALYTIC ACTIVITY: dtDP-6-deoxy-L-mannose + NADP(+) = dtDP-4-dehydro-6-deoxy-L-mannose + NADPH.			
CC	-1- COFACTOR: NADP.			
CC	-1- PATHWAY: dtDP-L-RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).			
CC				
DR	EMBL: X71970; CAA50768.1; -			
DR	PIR: S41535; S41535.			
DR	TIGR: TIGR01214; rmlD. 1			
KW	LiPolypolysaccharide biosynthesis; Oxidoreductase; NADP.			
FT	NP_BIND 7 NADP (POTENTIAL).			
SO	SEQUENCE 300 AA; 32757 MW; 0F9ABBI6AF83A9F9 CRC64;			
Query Match	Score 42; DB 1; Length 300;			
Best Local Similarity	58.38; Pred. No. 1.6;			
Matches	7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;			
QY	1 TSPLINHGOKL 12			
Db	122 TAPLVNNGGTRL 133			
RESULT 2				
MDR5_DROME	STANDARD;			
ID	MDR5_DROME	PRT;	1302 AA.	

AC 000748; 09VNRW3;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Multidrug resistance protein homolog 65 (P-glycoprotein 65).
 GN MDR5 OR CG10181.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID:7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Head;
 RX MEDLINE-91304385; PubMed-2072901;
 RT Wu C.-T., Budding M., Griffin M.S., Croop J.M.;
 RT "Isolation and characterization of Drosophila multidrug resistance
 RT gene homologs."; 11:3940-3948(1991).
 RL Mol. Cell. Biol. 11:3940-3948(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MS1127, and MS1125;
 RX MEDLINE-20467314; PubMed-11012721;
 RT Begun D.J., Whitely P.;
 RT "Genetics of alpha-amanitin resistance in a natural population of
 RT Drosophila melanogaster.";
 RL Heredity 85:184-190(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-20196006; PubMed-10731132;
 RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amandalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 RA Sultion G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wein K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Plannkoch C., Baldwin D.,
 RA Babbler R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Durbin K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegvam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matvei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshnell A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Sidenkhan K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sidenkhan I., Simpson M., Skupski M.P., Smith T.,
 RA Spher E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: M59077; AAA28680.1; -;
 DR EMBL: AF251287; AAF69147.1; -;
 DR EMBL: AF251286; AAF69146.1; -;
 DR EMBL: AF003563; AAF50669.1; -;
 DR PIR: B41249; B41249.
 DR FlyBase: FBgn0004513; Mdr65.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR01140; ABCtransportTM.
 DR Pfam: PF00005; ABC_tran; 2.
 DR Pfam: PF00664; ABC_membrane; 2.
 DR ProDom: PD000006; ABC_transport; 2.
 DR SMART: SM00382; AAA; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
 DR ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
 KW Multigene family.
 FT DOMAIN 1 48
 FT TRANSMEM 49 69
 FT DOMAIN 70 118
 FT TRANSMEM 119 147
 FT DOMAIN 148 194
 FT TRANSMEM 195 215
 FT DOMAIN 216 223
 FT TRANSMEM 224 242
 FT DOMAIN 243 302
 FT TRANSMEM 303 323
 FT DOMAIN 324 341
 FT TRANSMEM 342 362
 FT DOMAIN 363 731
 FT TRANSMEM 732 753
 FT DOMAIN 754 776
 FT TRANSMEM 777 798
 FT DOMAIN 799 852
 FT TRANSMEM 853 873
 FT DOMAIN 874 894
 FT TRANSMEM 895 896
 FT DOMAIN 897 956
 FT TRANSMEM 957 977
 FT DOMAIN 978 993
 FT TRANSMEM 994 1014
 FT DOMAIN 1015 1302
 FT TRANSMEM 1302 143784
 FT NP_BIND 440 447
 FT NP_BIND 1094 1101
 FT REPEAT 1 673
 FT REPEAT 674 1302
 FT CARBOHYD 103 103
 FT CONFLICT 369 369
 FT CONFLICT 678 678
 SQ SEQUENCE 1302 AA; 143784 MW; 39A7BCABFA31924A CRC64;
 Query Match 65.6%; Score 42; DB 1; Length 1302;
 Best Local Similarity 72.7%; Pred. No. 8.5;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 TSPUNHGOK 11
 Db 673 TSPUNHGOK 683
 RESULT 3
 YDEK_ECOLI STANDARD; FRT; 1325 AA.
 AC P32051; P76140; P77168;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical lipoprotein ydek precursor (ORF1).
 OS YDEK OR ORF1 OR B1510.
 GN Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RA MEDLINE=97426617; PubMed=9278503;
 RX Biatlner F.R., Plunkett G., Ili, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Mada C., Yamamoto Y., Horuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [3]
 RP SEQUENCE OF 595-1325 FROM N.A.
 RX MEDLINE=94100243; PubMed=8274505;
 RA Cartwright P.J., Tjoms M.W., Lithgow T., Hoef P.B., Hoogenraad N.J.;
 RT "An Escherichia coli gene showing a potential ancestral relationship
 to the genes for the mitochondrial import site proteins ISP42 and
 MOM38.";
 RL Biochim. Biophys. Acta 1153:345-347(1993).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 (Potential).
 CC -1- SIMILARITY: TO E.COLI YEAL.
 CC -1- SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT SITE PROTEINS
 ISP42 AND MOM38.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 FRAMESHIFT IN POSITION 653.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000248; AAC74583.1; -;
 DR EMBL: D90793; BAA15190.1; ALT_INIT.
 DR EMBL: D90794; BAA15197.1; ALT_INIT.
 DR EMBL: X73295; CAA51730.1; ALT_FRAME.
 DR PIR: S34315; S34315.
 DR Ecogene; EG11780; ydek.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 DR Hypothetical protein; Membrane; Lipoprotein; signal;
 KM Complete proteome.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 1325 HYPOTHETICAL LIPOPROTEIN YDEK.
 FT LIPID 19 13 N-ACYL DIGLYCERIDE (POTENTIAL).
 FT CONFLICT 884 N -> K (IN REF. 3).
 FT FT 1317 1317 M -> S (IN REF. 3).
 SQ SEQUENCE 1325 AA; 136514 MW; 26A3A066FA19AD7D CRC64;

Query Match 65.6%; Score 42; DB 1; Length 1325;
 Best Local Similarity 88.9%; Pred. No. 8.7;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNHNGOKL 12
 |||| ||||
 Db 991 LNHNGOKL 999
 RESULT 4
 ID HEX_ADEP3 STANDARD; PRT; 939 AA.
 AC Q9YTR8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hexon protein (Late protein 2).
 GN PIT.
 OS Porcine adenovirus type 3 (PAV-3).
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 OX NCBI_TaxID=35265;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=6618;
 RX MEDLINE=99058191; PubMed=9837805;
 RA Reddy P.S., Idamakanti N., Song J.Y., Lee J.B., Hyun B.H., Park J.H.,
 RA Cha S.H., Bae Y.T., Tikoo S.K., Babluk L.A.;
 RT "Nucleotide sequence and transcription map of porcine adenovirus type
 3.";
 RL Virology 251:414-426(1998).
 CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE STRUCTURAL PROTEINS IN THE
 VIRAL COAT AND IS SYNTHESIZED DURING LATE INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF083132; AAC99441.1; -;
 DR HSSP; P03277; IDHX.
 DR InterPro; IPR000736; Adeno_hexon.
 DR Pfam; PF01065; Adeno_hexon; 1.
 DR Pfam; PF03678; Adeno_hexon.C; 1.
 DR ProDom; PD002815; Adeno_hexon; 1.
 KW Cost protein; Hexon protein; Late protein.
 SQ SEQUENCE 939 AA; 106087 MW; 3B3B98EAC7C794EE CRC64;

Query Match 60.9%; Score 39; DB 1; Length 939;
 Best Local Similarity 77.8%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 SPLNHNQ 10
 || ||| ||
 Db 223 SPLNHNQ 231
 RESULT 5
 ID GLI4_HUMAN STANDARD; PRT; 223 AA.
 AC P10075;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE GLI4 protein (Kruessel-related zinc finger protein 4) (HKR4 protein)
 DE (Fragments).
 GN GLI4 OR HKR4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89096896; PubMed=2850480;

```

RA Ruppert J.M., Kitzler K.W., Wong A.J., Bigner S.H., Kao F.T.,
RA Law M.L., Senanez H.N., O'Brien S.J., Vogelstein B.;
RT "The Gli-Kruppel family of human genes.";
RL MOL. CELL. BIOL. 8:3104-3113(1988).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential)
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M20678; AAA35990.1; -.
DR EMBL: M20679; -. NOT_ANNOTATED_CDS.
DR PIR: F31201; F31201.
DR HSP: P08046; IAI6.
DR Genew: HGNC:4320; GLI4.
DR MIM: 165280; -.
DR InterPro: IPR000822; Znf_C2H2.
DR PRINTS: PR00048; ZINC_FINGER.
DR SMART: SM00355; Znf_C2H2; 6.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 6.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 6.
KW Zinc-finger; Metal-binding; DNA-binding; Nuclear protein; Repeat.
KW NUCLEAR PROTEIN; REPEAT.
FT DOMAIN 1 1
FT ZN_FING 40 >129 ZINC_FINGERS.
FT ZN_FING 40 62 C2H2-TYPE.
FT ZN_FING 68 90 C2H2-TYPE.
FT ZN_FING 96 118 C2H2-TYPE.
FT ZN_FING 124 >129 C2H2-TYPE.
FT ZN_FING 129 130 C2H2-TYPE.
FT ZN_FING 130 136 ZINC_FINGERS.
FT ZN_FING 136 136 C2H2-TYPE.
FT ZN_FING 142 164 C2H2-TYPE.
FT ZN_FING 170 192 C2H2-TYPE.
FT ZN_FING 198 220 C2H2-TYPE.
FT ZN_FING 220 220 C2H2-TYPE.
SQ SEQUENCE 223 AA; 25276 MW; 5DF24D508A04EA69 CRC64;

Query Match 59.4%; Score 38; DB 1; Length 223;
Best Local Similarity 63.6%; Pred. No. 6.7;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 TSPINHGOK 11
Db 122 TQHRINGEK 122

RESULT 6
2177_HUMAN STANDARD; PRT; 321 AA.
AC 013360;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein 177.
GN ZNF177.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RA MEDLINE-9629641; PubMed-8661005;
RA Baban S., Freeman J.D., Mager D.L.;
RT "Transcripts from a novel human KRA3 zinc finger gene contain spliced
RT Alu and endogenous retroviral segments.";
RL Genomics 33:463-472(1996).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).

```

```

CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRA3 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U37263; AAB09749.1; -.
DR HSP: P08047; ISP2.
DR Genew: HGNC:12966; ZNF177.
DR MIM: 601276; -.
DR InterPro: IPR001909; KRA3.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 7.
DR PRINTS: PR00048; ZINC_FINGER.
DR PRODOM: PD000003; Znf_C2H2; 7.
DR SMART: SM00349; KRA3; 1.
DR SMART: SM00355; Znf_C2H2; 7.
DR PROSITE: PS50805; KRA3; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
KW Nuclear protein; Repeat.
KW NUCLEAR PROTEIN; REPEAT.
FT DOMAIN 14 84 KRA3.
FT ZN_FING 124 314 ZINC_FINGERS.
FT ZN_FING 124 146 C2H2-TYPE.
FT ZN_FING 152 174 C2H2-TYPE.
FT ZN_FING 180 202 C2H2-TYPE.
FT ZN_FING 208 230 C2H2-TYPE.
FT ZN_FING 236 258 C2H2-TYPE.
FT ZN_FING 264 286 C2H2-TYPE.
FT ZN_FING 292 314 C2H2-TYPE.
SQ SEQUENCE 321 AA; 36473 MW; E325860C292DA17 CRC64;

Query Match 59.4%; Score 38; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGOKL 12
Db 313 IHNGOKL 319

RESULT 7
FBX3_HUMAN STANDARD; PRT; 471 AA.
ID FBX3_HUMAN
AC 09UK99; 09NUX2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE F-box only protein 3.
GN FBX3 OR FBX3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RA TISSUE-Placenta;
RA Iisoga T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "HEDD human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC [2]
CC SEQUENCE OF 6-415 FROM N.A.

```

```

RX MEDLINE-20003061; PubMed-10531037;
RA Winston J.T., Koepf D.M., Zhu C., Ellledge S.J., Harper J.W.;
RT "A family of mammalian F-box proteins.";
RL Curr. Biol. 9:1180-1182(1999).
CC -1- FUNCTION: Probably recognizes and binds to some phosphorylated
CC proteins and promotes their ubiquitination and degradation.
CC -1- SUBUNIT: PART OF A SCF (SKP1-CUL1IN-F-BOX) PROTEIN LIGASE COMPLEX
CC (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 F-BOX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AK01943; BAA91991.1; -
CC EMBL: AF176702; AAF03702.1; -
CC Genew: HGNC:13582; FBXO3.
CC InterPro: IPR001810; F-box.
CC Pfam: PF00646; F-box; 1.
CC SMART: SM00256; FBOX; 1.
CC PROSITE: PS50181; FBOX; 1.
CC Ubl conjugation.
CC KW DOMAIN 10 56 F-BOX.
CC FT DOMAIN 419 452 ASP/GLU-RICH (HIGHLY ACIDIC).
CC FT DOMAIN 453 456 POLY-ARG.
CC FT DOMAIN 463 466 POLY-ARG.
CC FT CONFLICT 164 164 T -> A (IN REF. 2).
CC FT CONFLICT 414 415 EM -> VS (IN REF. 2).
SQ SEQUENCE 471 AA; 54590 MW; F7AA88193E14E67E CRC64;

Query Match 59.4%; Score 38; DB 1; Length 471;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 IHNGOKL 12
Db 148 IHNGOKL 154

RESULT 8
PLOT_SCHPO
ID PLOT_SCHPO STANDARD; PRT; 683 AA.
AC P50528;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Serine/threonine-protein kinase plot (EC 2.7.1.1.-).
GN PLOT OR SPAC23C11.16
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCB1_TaxID=4896;
ON
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE-95262899; PubMed-7744248;
RA Ohkura H., Hagan I.M., Glover D.M.;
RT "The conserved Schizosaccharomycetes pombe kinase plot, required to
RT form a bipolar spindle, the actin ring, and septum, can drive septum
RT formation in G1 and G2 cells.";
RL Genes Dev. 9:1059-1073(1995).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE-21848401; PubMed-11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

```

```

RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidaigo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagals K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Slimmons M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grympeprez B.,
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Filtz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Beger P., Zimmermann W., Medler H., Manbuit R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Gallibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- FUNCTION: REQUIRED TO FORM A BIPOLAR SPINDLE, THE ACTIN RING AND
CC SEPTUM. FUNCTIONS UPSTREAM OF THE WHOLE SEPTUM FORMATION PATHWAY,
CC INCLUDING ACTIN RING FORMATION (REGULATED BY LATE SEPTATION GENES)
CC AND SEPTAL MATERIAL DEPOSITION (REGULATED BY EARLY SEPTATION
CC GENES). BEHAVES AS A "SEPTUM-PROMOTING FACTOR", AND COULD ALSO BE
CC INVOLVED IN INDUCING OTHER LATE EVENTS OF CELL DIVISION.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 2 POLO BOX DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X85758; CA59766.1; -
CC EMBL: Z98559; CAB1167.1; -
CC HSSP: 063450; IA06.
CC InterPro: IPR000719; Euk.pkinase.
CC InterPro: IPR000959; POLO_box.
CC InterPro: IPR002290; Ser_thr_pkinase.
CC Pfam: PF00069; pkinase; 1.
CC Pfam: PF00659; POLO_box; 2.
CC Prodom: PD000001; Euk.pkinase; 1.
CC SMART: SM00220; S_TKc; 1.
CC PROSITE: PS50078; POLO_BOX; 2.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC Transferrase; Serine/threonine-protein kinase; ATP-binding; Repeat.
CC KW DOMAIN 41 296 PROTEIN KINASE.
CC FT NE_BIND 47 296 ATP (BY SIMILARITY).
CC FT BINDING 69 69 ATP (BY SIMILARITY).
CC FT ACT_SITE 163 163 BY SIMILARITY.
CC FT DOMAIN 500 567 POLO BOX 1.
CC FT DOMAIN 604 670 POLO BOX 2.
SQ SEQUENCE 683 AA; 77301 MW; F11CD0EF9B913917 CRC64;

Query Match 59.4%; Score 38; DB 1; Length 683;
Best Local Similarity 54.5%; Pred. No. 23;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 TSPPTNHNGOK 11
Db 27 TSPPTNHNGOK 37

```

```

RESULT 9
SOHA_ECOLI
ID SOHA_ECOLI STANDARD: PRT; 111 AA.
AC P15373;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Htra suppressor protein (Protein p1f).
SO SOHA OR PRF OR B3129.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90170878; PubMed-2407727;
RA Baird L., Georgopoulos C.;
RT "Identification, cloning, and characterization of the Escherichia
RT coli soha gene, a suppressor of the htra (degP) null phenotype.";
RL J. Bacteriol. 172:1587-1594(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-90094216; PubMed-2152898;
RA Kline D.R., Phillips G.J., Silhavy T.J.;
RT "Increased expression of the bifunctional protein Prf suppresses
RT overproduction lethality associated with exported beta-galactosidase
RT hybrid proteins in Escherichia coli.";
RL J. Bacteriol. 172:185-192(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Blythe R., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mall B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -1- FUNCTION: PRF APPEARS TO BE A BIFUNCTIONAL PROTEIN, WITH THE
CC ABILITY TO REGULATE ITS OWN EXPRESSION AS WELL AS RELIEVE THE
CC EXPORT BLOCK IMPOSED BY HIGH-LEVEL SYNTHESIS OF THE LAMB-LACZ
CC HYBRID PROTEIN.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: M30178; AAA24638.1; -
CC EMBL: M32358; AAA24418.1; -
CC EMBL: U18997; AAA57932.1; -
CC EMBL: AE000394; AAC76163.1; -
CC PIR: A35137; A35137.
CC EcoGene: EG10955; soha.
CC Complete proteome.
SQ SEQUENCE 111 AA; 12359 MW; 5FC0D5FF43F75D8A CRC64;

Query Match 57.8%; Score 37; DB 1; Length 111;
Best Local Similarity 58.3%; Pred. No. 4.7;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 TSPLNHNGOKL 12
DB 84 TRPFNIQGGKTL 95

RESULT 10
REBD_SALTY
ID REBD_SALTY STANDARD: PRT; 299 AA.

```

```

AC P26392;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE dTDP-4-dehydrothamnose reductase (EC 1.1.1.133) (dTDP-4-keto-L-
DE thamnose reductase) (dTDP-6-deoxy-L-mannose dehydrogenase) (dTDP-L-
DE thamnose synthetase).
GN REBD OR STM2096.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-LT2;
RX MEDLINE-91260454; PubMed-1710759;
RA Jiang X.-M., Neal B., Santiago F., Lee S.J., Romana L.K., Reeves P.R.;
RT "Structure and sequence of the rfb (O antigen) gene cluster of
RT Salmonella serovar typhimurium (strain LT2)."
RL Mol. Microbiol. 5:695-713(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-LT2 / SGC1412 / ATCC 700720;
RX MEDLINE-21534948; PubMed-11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Portolillo S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grevell N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
CC -1- CATALYTIC ACTIVITY: dTDP-6-deoxy-L-mannose + NADP(+) = dTDP-4-
CC dehydro-6-deoxy-L-mannose + NADPH.
CC -1- PATHWAY: DNDP-L-RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X56793; CAA40116.1; -
CC EMBL: AB008792; AAL21000.1; -
CC PIR: S15300; S15300.
CC StyGene: SG10344; rfdD.
CC TIGRfams: TIGR01214; rmlD; 1.
CC Lipopolysaccharide biosynthesis; Oxidoreductase; NADP;
CC Complete proteome.
FT NP_BIND 7
SQ SEQUENCE 299 AA; 32554 MW; ABAAD476AF5ECDE7 CRC64;

Query Match 57.8%; Score 37; DB 1; Length 299;
Best Local Similarity 58.3%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 TSPLNHNGOKL 12
DB 122 TSPLNHYGKTKL 133

RESULT 11
GAG_BIV06
ID GAG_BIV06 STANDARD: PRT; 476 AA.
AC P19558;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GAG polypeptide (P53) [Contains: Core proteins P17, P26, P14].
GN GAG.

```



```

OS Bovine immunodeficiency virus (isolate 106) (BIV).
CC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxId=11658;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90223985; PubMed=2183467;
RA Garvey K.J., Oberste M.S., Elser J.E., Braun M.J., Gonda M.A.;
RT "Nucleotide sequence and genome organization of biologically active
RL proviruses of the bovine immunodeficiency-like virus.";
Virology 175:391-409(1990).
CC -1- SIMILARITY: CONTAINS 2 CCHC-TYPE ZINC FINGERS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M32691; NOT_ANNOTATED_CDS.
DR HIV: M32691; GAGSBI106.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; Zf-CCHC; 2.
DR Pfam: PF00607; Gag_p24; 1.
DR PRINTS: PR00939; C2HCZNFINGER.
DR SMART: SM00343; ZNF_C2HC; 2.
DR PROSITE: PS50158; ZF_CCHC; 2.
KW Core protein; Polyprotein; Zinc-finger.
FT CHAIN 1 133
FT CHAIN 134 360
FT CHAIN 361 476
FT ZN_FING 403 420
FT ZN_FING 421 438
FT SEQUENCE 476 AA; 53470 MW; 54DF23B95A80269 CRC64;

Query Match
Best Local Similarity 57.8%; Score 37; DB 1; Length 476;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 PLNHNQOK 11
Db 287 PINHOGPK 295
1:|||||
1:|||||

RESULT 12
GAG_BIV27 STANDARD; PRT; 476 AA.
AC P19559;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GAG polyprotein (P53) [Contains: Core proteins p17, p26, p14].
GN GAG.
OS Bovine immunodeficiency virus (isolate 127) (BIV).
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxId=11659;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90223985; PubMed=2183467;
RA Garvey K.J., Oberste M.S., Elser J.E., Braun M.J., Gonda M.A.;
RT "Nucleotide sequence and genome organization of biologically active
RL proviruses of the bovine immunodeficiency-like virus.";
Virology 175:391-409(1990).
CC -1- SIMILARITY: CONTAINS 2 CCHC-TYPE ZINC FINGERS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M32690; AAA91270.1; -.
DR PIR: A34742; FOLJRT.
DR HIV: M32690; GAGSBI127.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; Zf-CCHC; 2.
DR Pfam: PF00607; Gag_p24; 1.
DR PRINTS: PR00939; C2HCZNFINGER.
DR SMART: SM00343; ZNF_C2HC; 2.
DR PROSITE: PS50158; ZF_CCHC; 2.
KW Core protein; Polyprotein; Zinc-finger; Repeat.
FT CHAIN 1 133
FT CHAIN 134 360
FT CHAIN 361 476
FT ZN_FING 403 420
FT ZN_FING 421 438
FT SEQUENCE 476 AA; 53440 MW; FAA896BD684255FF CRC64;

Query Match
Best Local Similarity 57.8%; Score 37; DB 1; Length 476;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 PLNHNQOK 11
Db 287 PINHOGPK 295
1:|||||
1:|||||

RESULT 13
MINC_HELPJ STANDARD; PRT; 194 AA.
ID MINC_HELPJ
AC 092M51;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable septum site-determining protein minc.
GN MINC OR JHP0372.
OS Helicobacter pylori j99 (Campylobacter pylori j99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX Helicobacter.
OX NCBI_TaxId=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
Nature 397:176-180(1999).
RL CL
CC -1- FUNCTION: CELL DIVISION INHIBITOR THAT BLOCKS THE FORMATION OF
CC POLAR 2 RING SEPTUMS. RAPIDLY OSCILLATES BETWEEN THE POLES OF THE
CC CELL. TO DESTABILIZE FTSZ FILAMENTS THAT HAVE FORMED BEFORE THEY
CC MATURE INTO POLAR 2 RINGS. PREVENT FTSZ POLYMERIZATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: INTERACTS WITH MIND AND FTSZ (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE MINC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AE001472; AAD05953.1; -.
KW Cell division; Separation; Complete proteome.
SQ SEQUENCE 194 AA; 22302 MW; 33093FA8637D0FA8 CRC64;

```

Query Match 56.28; Score 36; DB 1; Length 194;
 Best Local Similarity 75.0%; Pred. No. 14;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 5 NINNGOKL 12
 DB 117 NINNGAKI 124

RESULT 14

MINC_HELPY STANDARD; PRT; 195 AA.
 AC 025693;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable septum site-determining protein minc.
 GN MINC OR HP1053.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 NCBI_TaxID=210;
 RN [1]

SEQUENCE FROM N.A.
 RC STRAIN-26693 / ATCC 700392;
 RX MEDLINE-97394467; PubMed-9252185;
 RA Tomb J.-F., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodok A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;

RT "The complete genome sequence of the gastric pathogen Helicobacter
 pylori.";
 RL Nature 388:539-547(1997).

CC -1- FUNCTION: CELL DIVISION INHIBITOR THAT BLOCKS THE FORMATION OF
 CC POLAR 2 RING SEPTUMS. RAPIDLY OSCILLATES BETWEEN THE POLES OF THE
 CC CELL TO DESTABILIZE FTSZ FILAMENTS THAT HAVE FORMED BEFORE THEY
 CC MATURE INTO POLAR 2 RINGS. PREVENT FTSZ POLYMERIZATION (BY
 CC SIMILARITY).

CC -1- SUBUNIT: INTERACTS WITH MIND AND FTSZ (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE MINC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE000613; AAD08105.1; ALT_INIT.
 DR TIGR: HPI053;

DR Cell division; Septation; Complete proteome.
 SQ SEQUENCE 195 AA; 22372 MW; 5C2DE7235B33C77B CRC64;

Query Match 56.28; Score 36; DB 1; Length 195;
 Best Local Similarity 75.0%; Pred. No. 14;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 5 NINNGOKL 12
 DB 118 NINNGAKI 125

RESULT 15

YGGJ_ECOLI STANDARD; PRT; 243 AA.
 AC P37912; P76647;
 DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein yggj.
 DE YGGJ OR B2946.
 GN YGGJ OR B2946.

OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_TaxID=562;
 RN [1]

SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).

RN [2]
 SEQUENCE OF 1-128 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-94156871; PubMed-8113204;
 RA Jekel M., Wackernagel W.;

RT "Location of the enda gene coding for endonuclease I on the physical
 map of the Escherichia coli K-12 chromosome.";
 RL J. Bacteriol. 176:1550-1551(1994).

RN [3]
 SEQUENCE OF 127-243 FROM N.A.
 RC STRAIN-B.
 RX MEDLINE-85087938; PubMed-6393055;
 RA Gushima H., Yasuda S., Soeda E., Yokota M., Kondo M., Kimura A.;

RT "Complete nucleotide sequence of the E. coli glutathione synthetase
 gsh-II.";
 RL Nucleic Acids Res. 12:9299-9307(1984).

RT [4]
 IDENTIFICATION.

RA Rudd K.E., Baum B.;

RL Unpublished observations (AUG-1994).
 CC -1- SIMILARITY: BELONGS TO THE UPF0088 FAMILY. STRONG, TO H. INFLUENZAE
 CC H10303.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U28377; AAA69113.1; ALT_INIT.
 CC EMBL: AE000377; AAC75983.1; ALT_INIT.

DR EMBL: X65169; -: NOT_ANNOTATED CDS.
 DR EMBL: X01666; -: NOT_ANNOTATED CDS.

DR EcoGene: EG12366; yggj.
 DR InterPro: IPR004382; Cons_hypoth46.

DR TIGR: TIGR00046; Cons_hypoth46; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 243 AA; 26978 MW; 5C4659F5B295E033 CRC64;

Query Match 56.28; Score 36; DB 1; Length 243;
 Best Local Similarity 77.8%; Pred. No. 17;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 SPLNINGO 10
 DB 76 SPLNHLGO 84

Search completed: January 3, 2003, 14:25:08
 Job time : 12 secs

Best Local Similarity 63.6%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 TSPLNHNQK 11
| | | | | | | |
Db 112 TQHLRIHNGEK 122

RESULT 13

T24465
hypothetical protein T04F8.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T24465

R:Lenard, N.
submitted to the EMBL Data Library, November 1995

A:Reference number: Z19895

A:Accession: T24465

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-324 <MIL>

A:Cross-references: EMBL:Z66565; PIDN:CAA91477.1; GSPDB:GN00028; CESP:T04F8.1

A:Experimental source: clone T04F8

C:Genetics:

A:Gene: CESP:T04F8.1

A:Map position: X

A:Introns: 29/3; 58/2; 172/3; 201/2; 244/1

C:Superfamily: Saccharomyces probable membrane protein YOR271C

Query Match 59.4%; Score 38; DB 2; Length 324;
Best Local Similarity 54.5%; Pred. No. 21;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 TSPLNHNQK 11
| | | | | | | |
Db 35 TNPLNLFHGER 45

RESULT 14

AD0833

probable membrane protein STY2859 [imported] - Salmonella enterica subsp. enterica serov

C:Species: Salmonella enterica subsp. enterica serovar Typhl

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C:Accession: AD0833

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; PMID:11677608

A:Accession: AD0833

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-406 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05851.1; PID:g16503826; GSPDB:GN00176

C:Genetics:

A:Gene: STY2859

Query Match 59.4%; Score 38; DB 2; Length 406;
Best Local Similarity 50.0%; Pred. No. 27;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 PLNHNQK 12
| | | | | | | |
Db 356 PFLNHNQRI 365

RESULT 15

T29991

hypothetical protein C43H6.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T29991

R:Le, T.T.

submitted to the EMBL Data Library, March 1996

A:Description: The sequence of C. elegans cosmid C43H6.

A:Reference number: Z20717

A:Accession: T29991

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-641 <LEF>

A:Cross-references: EMBL:U51999; PIDN:AAA96085.1; GSPDB:GN00028; CESP:C43H6.4

A:Experimental source: strain Bristol N2; clone C43H6

C:Genetics:

A:Gene: CESP:C43H6.4

A:Map position: X

A:Introns: 12/1; 63/2; 89/3; 143/1; 187/3; 268/3; 364/3; 403/3; 442/2; 487/2; 517/1

C:Superfamily: Caenorhabditis elegans hypothetical protein C43H6.4

Query Match 59.4%; Score 38; DB 2; Length 641;
Best Local Similarity 72.7%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 SPLNHNQK 12
| | | | | | | |
Db 510 SPLNHNQK 520

Search completed: January 3, 2003, 14:25:30
Job time: 16 secs

A:Molecule type: DNA
 A:Residues: 1-427 <BEV>
 A:Cross-references: EMBL:AL078464; GSPDB:GN00062; ATSP:F6G3.70
 A:Experimental source: cultivar Columbia; BAC clone F6G3
 C:Genetics:
 A:Gene: ATSP:F6G3.70
 A:Map position: 4

Query Match
 Best Local Similarity 64.1%; Score 41; DB 2; Length 427;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TSPLEIING 9
 DB 263 TTPLEIING 271

RESULT 9
 T36546
 hypothetical proline-rich protein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T36546
 R:Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, September 1997
 A:Reference number: Z21748
 A:Accession: T36546
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-331 <OLI>
 A:Cross-references: EMBL:Z99165; PIDN:CAB16268.1; GSPDB:GN00066; SPDB:SPAC2F3.14c
 A:Experimental source: strain 972h-; cosmid c2F3
 C:Genetics:
 A:Gene: SPDB:SPAC2F3.14c
 A:Map position: 1

Query Match
 Best Local Similarity 62.5%; Score 40; DB 2; Length 331;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 SPLNINHGOKL 12
 DB 271 SSLNINHGOKL 281

RESULT 9
 AF0114
 conserved hypothetical protein YPO0934 [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
 C:Accession: AF0114
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiliak, R.W.; Holden, M.T.G.; Prentice, M.B.
 deno-Farrage, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AF0114
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-243 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC8977.1; PID:g15979004; GSPDB:GN00175
 C:Genetics:
 A:Gene: YPO0934
 C:Superfamily: Haemophilus influenzae hypothetical protein HI0303

Query Match
 Best Local Similarity 60.9%; Score 39; DB 2; Length 243;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 SPLNINHGOKL 10
 DB 76 SPLNINHGOKL 84

RESULT 10
 T36580
 hypothetical protein SCH24.15c - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T36580
 R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, May 1999
 A:Reference number: Z21575
 A:Accession: T36580
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-563 <OLI>
 A:Cross-references: EMBL:AL049826; PIDN:CAB42719.1; GSPDB:GN00070; SCOEDB:SCH24.15c
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SCH24.15c

Query Match
 Best Local Similarity 60.9%; Score 39; DB 2; Length 563;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 TSPLEIINGOKL 12
 DB 57 TSPPELHSGHKL 68

RESULT 11
 F86635
 transporter yaJA [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 C:Accession: F86635
 R:Bohlooli, A.; Wincker, P.; Manger, S.; Jalllon, O.; Malarme, K.; Welssenbach, J.; Eh
 Genome Res. 11, 731-753, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
 A:Reference number: A86625; MUID:21235186; PMID:11337471
 A:Accession: F86635
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-150 <STO>
 A:Cross-references: GB:AE005176; PID:g12722930; PIDN:AAK04184.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: yaJA

Query Match
 Best Local Similarity 59.4%; Score 38; DB 2; Length 150;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 PLNINHGOKL 12
 DB 114 PLTVHNGENT 123

RESULT 12
 F31201
 GLI-related finger protein HKR4 - human (fragments)
 C:Species: Homo sapiens (man)
 C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 17-Mar-1994
 C:Accession: F31201
 R:Puppert, J.M.; Kinzler, K.W.; Wong, A.J.; Bigner, S.H.; Kao, F.T.; Law, M.L.; Seuan
 Mol. Cell. Biol. 8, 3104-3113, 1988
 A:Title: The GLI-Kruppel family of human genes.
 A:Reference number: A93103; MUID:8906896; PMID:2850480
 A:Accession: F31201
 A:Molecule type: DNA
 A:Residues: 1-223 <RUP>
 C:Keywords: DNA binding; zinc finger

Query Match
 Best Local Similarity 59.4%; Score 38; DB 2; Length 223;

Best Local Similarity 58.3%; Pred. No. 3.2;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 TSPININGOKL 12
|||||
Db 122 TAPLVNGGFKL 133

RESULT 3

A64905

ydek protein - Escherichia coli (strain K-12)

N:Alternate names: protein Y

C:Species: Escherichia coli

C>Date: 13-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C:Accession: A64905; 152440; S34315

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A>Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: A64905

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1325 <BLAT>

A:Cross-references: GB:AE000248; GB:U00096; MID:q1787783; PIDN:AACT4583.1; PID:q1787788;

A:Experimental source: strain K-12, substrain MG1655

R:Cartwright, P.; Timms, M.; Litgow, T.; Hoj, P.; Hoogenraad, N.

Biochim. Biophys. Acta 1153, 345-347, 1993

A>Title: An Escherichia coli gene showing a potential ancestral relationship to the gene

A:Reference number: 152440; MUID:94100243; PMID:8274505

A:Accession: 152440

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 689-883; 'K', 885-1316; 'S', 1318-1325 <RES>

A:Cross-references: EMBL:X73295; MID:9312392; PIDN:CAA51730.1; PID:9312393

A>Note: the difference in length is due to a frameshift error at pos 653

C:Genetics:

A:Gene: ydek

C:Function:

A:Description: probably involved in protein translocation apparatus

C:Keywords: nucleotide binding; P-loop

F:112-119/Region: nucleotide-binding motif A (P-loop)

Query Match 65.6%; Score 42; DB 2; Length 1325;
Best Local Similarity 88.9%; Pred. No. 18;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 LNIHNGOKL 12
|||||
Db 991 LNIHNGOKL 999

RESULT 4

E90893

hypothetical protein Ecs2117 [imported] - Escherichia coli (strain O157:H7, substrain R1

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: E90893

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: E90893

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1343 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA835540.1; PID:q13361583; GSPDB:GN00154

C:Experimental source: strain O157:H7, substrain R1MD 0509952

C:Genetics:

A:Gene: Ecs2117

Query Match 65.6%; Score 42; DB 2; Length 1343;

Best Local Similarity 88.9%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 LNIHNGOKL 12
|||||
Db 991 LNIHNGOKL 999

RESULT 5

D85724

hypothetical protein ydek [imported] - Escherichia coli (strain O157:H7, substrain ED

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: D85724

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

hler, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: D85724

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1343 <STO>

A:Cross-references: GB:AE005174; MID:912515159; PIDN:AAG56256.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: ydek

Query Match 65.6%; Score 42; DB 2; Length 1343;
Best Local Similarity 88.9%; Pred. No. 18;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 LNIHNGOKL 12
|||||
Db 991 LNIHNGOKL 999

RESULT 6

T31070

notch homolog - sea urchin (Lytechinus variegatus)

C:Species: Lytechinus variegatus (variegated urchin)

C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000

C:Accession: T31070

R:Sherwood, D.R.; McClay, D.R.

Development 124, 3363-3374, 1997

A>Title: Identification and localization of a sea urchin Notch homologue: insights in

A:Reference number: Z20966; MUID:97454256; PMID:9310331

A:Accession: T31070

A:Molecule type: mRNA

A:Residues: 1-2531 <SHF>

A:Cross-references: EMBL:AF000634; MID:92570350; PID:92570351; PIDN:AAB82088.1

C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

Query Match 65.6%; Score 42; DB 2; Length 2531;
Best Local Similarity 60.0%; Pred. No. 37;

Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPININGO 10
|||||
Db 2097 TSPMDMNGE 2106

RESULT 7

T08980

hypothetical protein F6G3.70 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999

C:Accession: T08980

R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.

Submitted to the Protein Sequence Database, May 1999

A:Reference number: Z16520

A:Accession: T08980

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:23:37 ; Search time 15 seconds

(Without alignments)
76.908 Million cell updates/sec

Title: US-09-899-376-1

Perfect score: 64

Sequence: 1 TSPLNHNQOKL 12

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Capext 0.5

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: PIR_73:*

2: PIR1:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	68.8	1302	2 B41249	multidrug resistant
2	42	65.6	300	2 S41535	rfbc protein - Shi
3	42	65.6	1325	2 A64905	yeak protein - Esc
4	42	65.6	1343	2 E90893	hypothetical prote
5	42	65.6	1343	2 D85724	hypothetical prote
6	42	65.6	2531	2 T31070	notch homolog - se
7	41	64.1	427	2 T08980	hypothetical prote
8	40	62.5	331	2 T38546	hypothetical proti
9	39	60.9	243	2 AF0114	conserved hypothet
10	39	60.9	353	2 T36580	hypothetical prote
11	38	59.4	150	2 F86635	transporter yajA l
12	38	59.4	223	2 F31201	GLI-related finger
13	38	59.4	324	2 T24465	hypothetical prote
14	38	59.4	406	2 AD0833	probable membrane
15	38	59.4	641	2 T29991	hypothetical prote
16	38	59.4	683	2 T38254	serine/threonine-s
17	38	59.4	693	2 C84495	hypothetical prote
18	37	57.8	111	2 A35137	suppressor protein
19	37	57.8	111	2 G91129	suppressor protein
20	37	57.8	111	2 G85974	probable proteins
21	37	57.8	240	2 F83592	conserved hypothet
22	37	57.8	248	2 S23348	hypothetical prote
23	37	57.8	299	2 S15300	ribA protein - Sal
24	37	57.8	299	2 AE0767	drdp-4-dehydroxam
25	37	57.8	476	1 FOLJBT	gag polypeptide -
26	37	57.8	587	2 S58319	hypothetical prote
27	37	57.8	838	2 T09054	capsaicin receptor
28	36	56.2	169	2 G83075	type 4 fibrillar bl
29	36	56.2	194	2 H71939	hypothetical prote

30	36	56.2	217	2 E64651	hypothetical prote
31	36	56.2	243	2 AF0877	conserved hypothet
32	36	56.2	252	2 A63080	hypothetical prote
33	36	56.2	252	2 F91106	hypothetical prote
34	36	56.2	252	2 A85952	hypothetical prote
35	36	56.2	253	2 G84977	hypothetical prote
36	36	56.2	265	1 I40649	hypothetical prote
37	36	56.2	307	2 S75208	dnaf protein - Syn
38	36	56.2	372	2 T25717	hypothetical prote
39	36	56.2	454	2 A97048	similar to phospho
40	36	56.2	605	2 H71303	hypothetical prote
41	36	56.2	630	2 S70909	transferrin-bindin
42	36	56.2	817	2 T21336	hypothetical prote
43	36	56.2	1005	2 T31333	beta-galactosidase
44	36	56.2	1474	2 B85188	retrotransposon 11
45	35	54.7	15	2 A60929	dichloromethane de

ALIGNMENTS

RESULT 1

B41249
multidrug resistance protein homolog Mdr65 - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 03-Apr-1992 #sequence_revision 12-Jun-1992 #text_change 02-Feb-2001
C:Accession: B41249
R:Wu, C.T.: Budding, M.: Griffin, M.S.: Croop, J.M.
Mol. Cell. Biol. 11, 3940-3948, 1991
A:Title: Isolation and characterization of *Drosophila* multidrug resistance gene homol
A:Reference number: A41249; MUID:91304385; PMID:2072901
A:Accession: B41249
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1302 <MUA>
A:Cross-references: GB:M59077; NID:g157874; PIDN:AAA28680.1; PID:g157875
C:Genetics:
A:Gene: FlyBase:Mdr65
A:Cross-references: FlyBase:FBgn0004513
C:Superfamily: unassigned ATP-binding cassette proteins: ATP-binding cassette homolog
C:Keywords: ATP; nucleotide binding; P-loop; transmembrane protein
F:423-617/Domain: ATP-binding cassette homolog <ABC1>
F:440-447/Region: nucleotide-binding motif A (P-loop)
F:1077-1274/Domain: ATP-binding cassette homolog <ABC2>
F:1094-1101/Region: nucleotide-binding motif A (P-loop)

Query Match

Best local Similarity 68.8%; Score 44; DB 2; Length 1302;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TSPLNHNQOK 11
Db 673 TSPLNHNQOK 683

RESULT 2

S41535
rfbc protein - *Shigella flexneri*
C:Species: *Shigella flexneri*
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 22-Jun-1999
C:Accession: S41535
R:Macpherson, D.F.: Manning, P.A.: Morona, R.
Mol. Microbiol. 11, 281-292, 1994
A:Title: Characterization of the drdp-rhamnose biosynthetic genes encoded in the rfbc
A:Reference number: S41533; MUID:94224146; PMID:8170390
A:Accession: S41535
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-300 <MAC>
A:Cross-references: EMBL:X71970; NID:g506557; PIDN:CMA50768.1; PID:g454899
C:Superfamily: drdp-dihydrostreptose synthase
Query Match 65.6%; Score 42; DB 2; Length 300;

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:22:12 ; Search time 35 Seconds
(without alignments)
45.686 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 64
Sequence: 1 TSPLNINHGOKL 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002:.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64	100.0	12	23	AA050916
2	64	100.0	19	23	AA050917
3	64	100.0	19	23	AA050918
4	42	65.6	804	22	ABG28800
5	42	65.6	804	22	ABG29023
6	42	65.6	807	22	ABG28665
7	42	65.6	1023	22	ABG28668
8	42	65.6	1302	22	ABG64919
9	42	65.6	1325	22	ABG98256
10	40	62.5	61	22	AB041576

11	40	62.5	61	22	AB025416	Protein #7415 enco
12	40	62.5	61	22	AA062447	Human brain expres
13	40	62.5	61	22	AA075256	Human bone marrow
14	40	62.5	61	22	AA035368	Peptide #9405 enco
15	40	62.5	61	22	AB044904	Human peptide enco
16	39	60.9	392	22	AB094887	Human protein sequ
17	39	60.9	2176	22	AB061386	Drosophila melanog
18	38	59.4	82	23	AB049299	Human ORFX protein
19	38	59.4	150	23	AB053887	Lactococcus lactis
20	38	59.4	247	21	AA083081	F-box protein FBP-
21	38	59.4	408	22	AA048304	Human ZF23 protein
22	38	59.4	444	22	AA064526	P. hybrida pollen-
23	38	59.4	444	22	AA064874	Pollen fertility r
24	38	59.4	444	22	AA062447	Zinc finger transc
25	38	59.4	444	22	AA062620	Male sterile plant
26	38	59.4	467	23	AB038350	Staphylococcus epi
27	38	59.4	471	21	AA090287	Human peptidase, H
28	38	59.4	471	23	AA093481	Human protein sequ
29	38	59.4	471	23	AA047764	F-cassette structu
30	38	59.4	472	22	AA082404	S. epidermidis ope
31	37	57.8	112	22	AB030012	Novel human diagno
32	37	57.8	179	23	AA049525	B. mori telomere c
33	37	57.8	240	22	AA035570	Pseudomonas aerugi
34	37	57.8	299	20	AA088336	Salmonella enteric
35	37	57.8	653	23	AA049524	B. mori telomere c
36	37	57.8	838	20	AA065555	Rat capsalcin rece
37	37	57.8	838	20	AA097899	Rat VRI capsalcin
38	37	57.8	838	22	AA061228	Rattus vanilloid r
39	37	57.8	1695	22	AB061826	Drosophila melanog
40	37	57.8	1711	22	AB061826	Drosophila melanog
41	36	56.2	194	22	AA035907	Helicobacter pylori
42	36	56.2	196	22	AA036067	Klebsiella pneumon
43	36	56.2	217	19	AA098678	H. pylori GHP0 572
44	36	56.2	252	21	AA011370	E. coli Y66F prote
45	36	56.2	252	22	AA034703	E. coli cellular p

ALIGNMENTS

RESULT 1
AA050916
ID AA050916 standard; Peptide: 12 AA.
XX
AC AA050916;
XX
DT 07-MAY-2002 (first entry)
XX
DE Tumour infiltrating peptide HN-1.
XX
KW Tumour infiltrating peptide; HN-1; head and neck cancer; HNSCC;
KW breast cancer; drug delivery; diagnosis; imaging; gene therapy.
XX
OS Synthetic.
XX
PN WO200202147-A2.
XX
PD 10-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-US21518.
XX
PR 30-JUN-2000; 2000US-215491P.
XX
PA (TEXTA) UNIV TEXAS SYSTEM.
XX
PI Clayman G, Hong FD;
XX
DR WPI; 2002-195737/25.
XX
PT Peptide internalised by a tumour cell useful for targeted delivery of
PT anticancer drugs -
XX
PS Claim 2; Page 71; 104pp; English.

XX The present sequence is that of a synthetic peptide, termed HN-1,
CC that is specifically internalised by human head and neck squamous
CC carcinoma cells (HNSCC), or certain other solid tumour tissue cells,
CC such as breast cancer cells. HN-1 was identified by screening a
CC phage M13 peptide library displaying over 10 power 9 peptides. The
CC screening method was based on the ability of HNSCC line MDAMB231
CC cells to uptake peptides by endocytosis at 37 degrees C. The cells
CC exhibited an approximately 10-fold greater internalisation
CC potential for HN-1 than normal human fibroblasts. The peptide
CC localised in cytoplasm after entry. In vivo, i.v. injected HN-1
CC peptide localised to HNSCC xenograft formed in nude mice. Claimed
CC compositions comprise HN-1 and a chemotherapeutic, cytotoxic,
CC apoptotic or DNA-damaging drug, such as taxol. The compositions
CC are used in claimed methods for killing a tumour cell, especially
CC an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal
CC cavity, larynx, thyroid, parathyroid, salivary gland, face or neck
CC skin or cervical lymph node cell), breast cancer cell or other
CC solid tumour cell. Also claimed are: methods for detecting cancer
CC by labeling HN-1 with a detectable label; tumour detection, tumour
CC imaging and tumour treating kits; methods for killing tumour cells
CC in which a composition comprising an antitumour compound conjugated
CC to HN-1 is administered with radiotherapy, chemotherapy, surgery or
CC a gene therapy composition; and a method for the isolating of an
CC internalising peptide by phage display library screening. The
CC peptide provides the necessary dose of a drug specifically to the
CC tumours, avoiding harmful side effects on other cells. The peptide
CC is non-toxic, non-immunogenic, stable in vivo, protects its cargo
CC during transit, and accumulates in a tumour within 48 hours.

XX Sequence 12 AA:

Query Match 100.0%; Score 64; DB 23; Length 12;

Best Local Similarity 100.0%; Pred. No. 1e-05;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLNHNGOKL 12

Db 1 TSPLNHNGOKL 12

RESULT 12

AAM50917 standard; Peptide: 19 AA.

XX AAM50917:

XX 07-MAY-2002 (first entry)

XX Tumour infiltrating peptide HN-2.

XX Tumour infiltrating peptide; HN-2; head and neck cancer; HNSCC;
KW breast cancer; drug delivery; diagnosis; imaging; gene therapy.

XX Synthetic.

XX Key Location/Qualifiers

FT Peptide 4..15 /note="corresponds to HN-1"

XX WO200202147-A2.

XX 10-JAN-2002.

XX 02-JUL-2001; 2001WO-US21518.

XX 30-JUN-2000; 2000US-215491P.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Clayman G, Hong FD;

XX PI

XX DR

XX Peptide internalised by a tumour cell useful for targeted delivery of
PT anticancer drugs -
XX Example 2; Page 76; 104pp; English.

XX The present sequence is that of a synthetic peptide, termed HN-2,
CC which is based on the claimed tumour infiltrating peptide HN-1 (see
CC AAM50916) with additional N- and C-terminal amino acid residues.
CC HN-1 is specifically internalised by human head and neck squamous
CC carcinoma cells (HNSCC) and certain other solid tumour tissue cells,
CC such as breast cancer cells. The additional amino acid residues of
CC HN-2 did not inhibit cell internalisation of the peptide. HN-1
CC internalisation is position-independent but sequence-dependent.
CC Claimed compositions comprise HN-1 and a chemotherapeutic, cytotoxic,
CC apoptotic or DNA-damaging drug, such as taxol. The compositions
CC are used in claimed methods for killing a tumour cell, especially
CC an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal
CC cavity, larynx, thyroid, parathyroid, salivary gland, face or neck
CC skin or cervical lymph node cell), breast cancer cell or other
CC solid tumour cell. Also claimed are: methods for detecting cancer
CC by labeling HN-1 with a detectable label; tumour detection, tumour
CC imaging and tumour treating kits; methods for killing tumour cells
CC in which a composition comprising an antitumour compound conjugated
CC to HN-1 is administered with radiotherapy, chemotherapy, surgery or
CC a gene therapy composition; and a method for the isolating of an
CC internalising peptide by phage display library screening. The
CC peptide provides the necessary dose of a drug specifically to the
CC tumours, avoiding harmful side effects on other cells. The peptide
CC is non-toxic, non-immunogenic, stable in vivo, protects its cargo
CC during transit, and accumulates in a tumour within 48 hours.

XX Sequence 19 AA:

Query Match 100.0%; Score 64; DB 23; Length 19;

Best Local Similarity 100.0%; Pred. No. 1.8e-05;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLNHNGOKL 12

Db 4 TSPLNHNGOKL 15

RESULT 3

AAM50918 standard; Peptide: 19 AA.

XX AAM50918:

XX 07-MAY-2002 (first entry)

XX Tumour infiltrating peptide HN-3.

XX Tumour infiltrating peptide; HN-3; head and neck cancer; HNSCC;
KW breast cancer; drug delivery; diagnosis; imaging; gene therapy.

XX Synthetic.

XX Key Location/Qualifiers

FT Peptide 8..19 /note="corresponds to HN-1"

XX WO200202147-A2.

XX 10-JAN-2002.

XX 02-JUL-2001; 2001WO-US21518.

XX 30-JUN-2000; 2000US-215491P.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Clayman G, Hong FD;

XX PI

XX WPI: 2002-195737/25.
DR
XX Peptide internalised by a tumour cell useful for targeted delivery of
PT anticancer drugs -
XX
XX Example 2; Page 76; 104pp; English.
PS
XX The present sequence is that of a synthetic peptide, termed HN-3,
CC which is based on the claimed tumour infiltrating peptide HN-1 (see
CC AAM50916) with additional N-terminal amino acid residues. HN-1 is
CC specifically internalised by human head and neck squamous carcinoma
CC cells (HNSCC) and certain other solid tumour tissue cells, such as
CC breast cancer cells. The additional N-terminal amino acid residues
CC of HN-3 did not inhibit cell internalisation of the peptide: HN-1
CC internalisation is position-independent but sequence-dependent.
CC Claimed compositions comprise HN-1 and a chemotherapeutic, cytotoxic,
CC apoptotic or DNA-damaging drug, such as taxol. The compositions
CC are used in claimed methods for killing a tumour cell, especially
CC an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal
CC cavity, larynx, thyroid, parathyroid, salivary gland, face or neck
CC skin or cervical lymph node cell), breast cancer cell or other
CC solid tumour cell. Also claimed are: methods for detecting cancer
CC by labeling HN-1 with a detectable label; tumour detection, tumour
CC imaging and tumour treating kits; methods for killing tumour cells
CC in which a composition comprising an antitumour compound conjugated
CC to HN-1 is administered with radiotherapy, chemotherapy, surgery or
CC a gene therapy composition; and a method for the isolating of an
CC internalising peptide by phage display library screening. The
CC peptide provides the necessary dose of a drug specifically to the
CC tumours, avoiding harmful side effects on other cells. The peptide
CC is non-toxic, non-immunogenic, stable in vivo, protects its cargo
CC during transit, and accumulates in a tumour within 48 hours.
XX
SQ Sequence 19 AA;
OY Query Match 100.0%; Score 64; DB 23; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TSPLNTHNGOKL 12
Db 8 TSPLNTHNGOKL 19
RESULT 4
ABG28800
ID ABG28800 standard; Protein: 804 AA.
XX
XX ABG28800;
AC
XX
XX 18-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #28791.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PA
XX (HYSE-) HYSEQ INC.
PI
XX Drmanac RT, Liu C, Tang YT;
XX

DR WPI: 2001-639362/73.
DR
XX N-PSDB; AAS92987.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 59159; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polynuclease chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 804 AA;
OY Query Match 65.6%; Score 42; DB 22; Length 804;
Best Local Similarity 88.9%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 4 LNIHNGOKL 12
Db 325 LNIHNGOKL 333
RESULT 5
ABG29023
ID ABG29023 standard; Protein: 804 AA.
XX
XX ABG29023;
AC
XX
XX 18-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #29014.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PA
XX (HYSE-) HYSEQ INC.
PI
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR
XX N-PSDB; AAS93210.
DR

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20: SEQ ID No 59382; 103bp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 804 AA:
Query Match 65.6%; Score 42; DB 22; Length 804;
Best Local Similarity 88.9%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 LNIHNGQKL 12
|||||
Db 325 LNIHNGQKL 333
RESULT 6
ABG28665
ID ABG28665 standard; Protein: 807 AA.
XX
AC ABG28665;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #28656.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
OS
PN WO200175067-A2.
PN
PD 11-OCT-2001.
PD
PF 30-MAR-2001; 2001WO-US08631.
PF
PR 31-MAR-2000; 2000US-0540217.
PR
PR 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR
DR N-PSDB; AAS92852.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20: SEQ ID No 59024; 103bp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 807 AA:
Query Match 65.6%; Score 42; DB 22; Length 807;
Best Local Similarity 88.9%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 LNIHNGQKL 12
|||||
Db 328 LNIHNGQKL 336
RESULT 7
ABG28668
ID ABG28668 standard; Protein: 1023 AA.
XX
AC ABG28668;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #28659.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
OS
PN WO200175067-A2.
PN
PD 11-OCT-2001.
PD
PF 30-MAR-2001; 2001WO-US08631.
PF
PR 31-MAR-2000; 2000US-0540217.
PR
PR 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR
DR N-PSDB; AAS92855.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

PT biodiversity -
XX
PS Claim 20: SEQ ID No 59027; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1023 AA;

Query Match 65.6%; Score 42; DB 22; Length 1023;
Best Local Similarity 88.9%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 LNIHNGOKL 12
|||||
Db 439 LNIHNGOKL 447

RESULT 8
ABB64919
ID ABB64919 standard; Protein; 1302 AA.

AC ABB64919;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 21549.

KW Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

OS

PN WO200171042-A2.

PD 27-SEP-2001.

PE 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PS 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

DR WPI: 2001-656860/75.

XX N-PSDB: ABL09022.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PS Disclosure: SEQ ID NO 21549; 21pp + Sequence Listing; English.

XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABJ30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1302 AA;

Query Match 65.6%; Score 42; DB 22; Length 1302;
Best Local Similarity 72.7%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TSPLNHNGOK 11
|||||
Db 673 TSPLNHNGOK 683

RESULT 9
AAG98256
ID AAG98256 standard; Protein; 1325 AA.

AC AAG98256;

DT 21-SEP-2001 (first entry)

XX Escherichia coli protein sequence SEQ ID NO:304.

KW Escherichia coli; identification; proliferation; microorganism;
KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;

KW bacterial growth inhibition.

XX Escherichia coli.

PN WO200148209-A2.

PD 05-JUL-2001.

PE 19-DEC-2000; 2000WO-US34419.

PR 23-DEC-1999; 99US-0173005.

PA (ELIT-) ELITRA PHARM INC.

PI Forsyth RA, Ohlsen KL, Zyskind JW;

DR WPI: 2001-457376/49.

XX N-PSDB: AAB81312.

PT Novel nucleic acids encoding proteins required for Escherichia coli
PT proliferation, useful for screening for antimicrobial agents -

PS Claim 19; Page 445-448; 596pp; English.

XX The present invention describes a purified or isolated nucleic acid
CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC given in AAB81202 to AAB81294, where expression of the nucleic acid in a
CC microorganism is capable of inhibiting proliferation of a microorganism.
CC (I) have antibacterial and antibiotic activities, and can be used in
CC gene therapy. Expression of (I) in a microorganism inhibits proliferation
CC of the microorganism, and the manufactured antibiotic is useful for
CC reducing the activity or level of a gene product required for
CC proliferation of a microorganism in a subject, specifically humans. The
CC nucleic acids that inhibit bacterial growth or proliferation can be used
CC as antisense therapeutics for killing bacteria. In addition to
CC therapeutic applications, the nucleic acid sequences complementary to
CC sequences required for proliferation can be used as diagnostic tools.

CC For example, nucleic acid probes complementary to proliferation-regulated
CC sequences that are specific for particular species of microorganisms can
CC be used as probes to identify particular microorganism species in
CC clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
CC proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
CC represent oligonucleotides, which are used in the exemplification of the
CC present invention.

XX
XX
SQ Sequence 1325 AA;

Query Match 65.6%; Score 42; DB 22; Length 1325;

Best Local Similarity 88.9%; Pred. No. 45;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNINHGOKL 12

DB 991 LNINHGOKL 999

RESULT 10

ABBA1576

ID ABBA1576 standard; Peptide; 61 AA.

XX
AC ABBA1576;

XX
DT 04-FEB-2002 (first entry)

XX
DE Peptide #9082 encoded by human foetal liver single exon probe.

XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX
OS Homo sapiens.

XX
PN MO200157277-A2.

XX
PD 09-AUG-2001.

XX
PF 30-JAN-2001; 2001WO-US00669.

XX
PR 04-FEB-2000; 2000US-0180312.

XX
PR 26-MAY-2000; 2000US-0207456.

XX
PR 30-JUN-2000; 2000US-0608408.

XX
PR 03-AUG-2000; 2000US-0632366.

XX
PR 21-SEP-2000; 2000US-0234687.

XX
PR 27-SEP-2000; 2000US-0236359.

XX
PR 04-OCT-2000; 2000GB-0024263.

XX
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX
DR WPI; 2001-483447/52.

XX
PT Human genome-derived single exon nucleic acid probes useful for

XX
PT analyzing gene expression in human foetal liver.

XX
PS Claim 27; SEQ ID NO 34211; 639pp + sequence listing; English.

XX
CC The invention relates to a single exon nucleic acid probe for

XX
CC measuring human gene expression in a sample derived from human foetal

XX
CC liver. The single exon nucleic acid probes may be used for predicting,

XX
CC measuring and displaying gene expression in samples derived from human

XX
CC foetal liver. The present sequence is a peptide encoded by a single exon

XX
CC nucleic acid probe of the invention.

XX
CC Note: The sequence data for this patent did not form part of the

XX
CC printed specification, but was obtained in electronic format directly

XX
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX
SQ Sequence 61 AA;

Query Match 62.5%; Score 40; DB 22; Length 61;

Best Local Similarity 72.7%; Pred. No. 3.1;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPKLNHGOK 11

DB 35 TSPKLNHGOK 45

RESULT 11

ABBA25416

ID ABBA25416 standard; Protein; 61 AA.

XX
AC ABBA25416;

XX
DT 23-JAN-2002 (first entry)

XX
DE Protein #7415 encoded by probe for measuring heart cell gene expression.

XX
KW Human; gene expression; heart; microarray; vascular system;

XX
KW cardiovascular disease; hypertension; cardiac arrhythmia;

XX
KW congenital heart disease.

XX
OS Homo sapiens.

XX
PN MO200157274-A2.

XX
PD 09-AUG-2001.

XX
PF 30-JAN-2001; 2001WO-US00666.

XX
PR 04-FEB-2000; 2000US-0180312.

XX
PR 26-MAY-2000; 2000US-0207456.

XX
PR 30-JUN-2000; 2000US-0608408.

XX
PR 03-AUG-2000; 2000US-0632366.

XX
PR 21-SEP-2000; 2000US-0234687.

XX
PR 27-SEP-2000; 2000US-0236359.

XX
PR 04-OCT-2000; 2000GB-0024263.

XX
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX
DR WPI; 2001-488899/53.

XX
PT single exon nucleic acid probes for analyzing gene expression in human

XX
PT hearts.

XX
PS Claim 15; SEQ ID NO 27186; 530pp; English.

XX
CC The present invention relates to single exon nucleic acid probes for

XX
CC measuring human gene expression in a sample derived from human heart (see

XX
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such

XX
CC probe. The probes may be used for predicting, measuring and displaying

XX
CC gene expression in samples derived from the human heart via microarrays.

XX
CC By measuring gene expression, the probes are useful for predicting,

XX
CC diagnosing, grading, staging, monitoring and prognosing diseases of the

XX
CC human heart and vascular system e.g. cardiovascular disease,

XX
CC hypertension, cardiac arrhythmias and congenital heart disease.

XX
CC Note: The sequence data for this patent did not form part of the printed

XX
CC specification, but was obtained in electronic format directly from WIPO

XX
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX
SQ Sequence 61 AA;

Query Match 62.5%; Score 40; DB 22; Length 61;

Best Local Similarity 72.7%; Pred. No. 3.1;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPKLNHGOK 11

DB 35 TSPKLNHGOK 45

RESULT 12

AAAB62447

ID AAM62447 standard; Protein: 61 AA.
 XX
 AC AAM62447;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 34552.
 XX
 KM Human; brain expressed exon; gene expression analysis; probe;
 KM microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KM epilepsy; cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157275-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00667.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483446/52.
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX
 PS Example 4; SEQ ID NO: 34552; 650bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.
 XX
 SQ Sequence 61 AA;
 XX

Query Match 62.5%; Score 40; DB 22; Length 61;
 Best Local Similarity 72.7%; Pred. No. 3.1;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 TSPKIHNGOK 11
 DB 35 TSPKIHNGOK 45
 DB

RESULT 13
 ID AAM75256 standard; Protein: 61 AA.
 XX
 AC AAM75256;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 35562.
 XX
 KM Human; bone marrow expressed exon; gene expression analysis; probe;
 KM microarray; cancer; leukemia; lymphoma; myeloma.
 XX
 OS Homo sapiens.
 XX

PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00668.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 PS Example 4; SEQ ID NO: 35562; 658bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.
 XX
 SQ Sequence 61 AA;
 XX

Query Match 62.5%; Score 40; DB 22; Length 61;
 Best Local Similarity 72.7%; Pred. No. 3.1;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 TSPKIHNGOK 11
 DB 35 TSPKIHNGOK 45
 DB

RESULT 14
 ID AAM35368 standard; Protein: 61 AA.
 XX
 AC AAM35368;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Peptide #9405 encoded by probe for measuring placental gene expression.
 XX
 KM Probe; microarray; human; placenta; antenatal diagnosis;
 KM genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX

XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488897/53.
 DR Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 27; SEQ ID NO 35637; 654bp; English.
 CC The present invention relates to single exon nucleic acid probes (SENP:
 CC see A013315-A157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX
 SQ Sequence 61 AA:
 Query Match 62.5%; Score 40; DB 22; Length 61;
 Best local Similarity 72.7%; Pred. No. 3.1;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 OY 1 TSPUNHNGOK 11
 111 : 1 1111
 Db 35 TSPKHNKGOK 45
 RESULT 15
 ABG44904
 ID ABG44904 standard; Peptide: 61 AA.
 XX ABG44904;
 AC
 XX 19-AUG-2002 (first entry)
 DT
 XX Human peptide encoded by genome-derived single exon probe SEQ ID 34569.
 DE
 XX Human: single exon probe; asthma; lung cancer; COPD; ILD;
 XX chronic obstructive pulmonary disease; interstitial lung disease;
 XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
 XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 XX Herimansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 XX primary ciliary dyskinesia; pulmonary hypertension;
 XX hyaline membrane disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-224687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2002-114183/15.
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX

PS Claim 27; SEQ ID NO 34569; 634bp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes: the novel set of probes which hybridize at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a peptide/protein
 CC encoded by a single exon probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 61 AA:
 Query Match 62.5%; Score 40; DB 23; Length 61;
 Best local Similarity 72.7%; Pred. No. 3.1;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 OY 1 TSPUNHNGOK 11
 111 : 1 1111
 Db 35 TSPKHNKGOK 45

Search completed: January 3, 2003, 14:24:51
 Job time : 37 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:25:38 : Search time 35 Seconds
(without alignments)
45.686 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 12
Sequence: 1 TSPLNHNGOKL 12

Scoring table: 101002
Gapop 60.0 , Gapext 60.0

Searched: 908470 segs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq_101002: *
1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT: *
2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT: *
3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT: *
4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT: *
5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT: *
6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT: *
7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT: *
8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT: *
9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT: *
10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT: *
11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT: *
12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT: *
13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT: *
14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT: *
15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT: *
16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT: *
17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT: *
18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT: *
19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT: *
20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT: *
21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT: *
22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT: *
23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12	100.0	12	23	AA50916 Tumour infiltratin
2	12	100.0	19	23	AA50917 Tumour infiltratin
3	12	100.0	19	23	AA50918 Tumour infiltratin
4	7	58.3	247	21	AA583081 F-box protein FBP-
5	7	58.3	408	22	AA58304 Human ZP23 protein
6	7	58.3	471	21	AA590287 Human peptidase, H
7	7	58.3	471	22	AA593481 Human protein sequ
8	7	58.3	471	23	AA547764 F-cassette structu
9	6	50.0	109	23	ABP38678 Staphylococcus epi
10	6	50.0	130	21	AA537863 Arabidopsis thalia

11	6	50.0	162	22	AB524051 Novel human diagno
12	6	50.0	467	23	ABP38390 Staphylococcus epi
13	6	50.0	472	22	AA582404 S. epidermidis ope
14	5	41.7	20	16	AA513187 Fragment of p53 bl
15	5	41.7	20	19	AA57358 Human WBP1 immunog
16	5	41.7	20	22	AA566642 Rabbit tissue fact
17	5	41.7	39	21	AA544615 Human secreted pro
18	5	41.7	45	20	AA567843 Human secreted pro
19	5	41.7	45	20	AA563358 Human breast cance
20	5	41.7	52	19	AA539901 pig plasma membran
21	5	41.7	54	22	AA564431 Human immune/haema
22	5	41.7	59	19	AA57369 Human WBP1 polyval
23	5	41.7	62	22	AA50564 Propionibacterium
24	5	41.7	63	22	AB567816 Drosophila melanog
25	5	41.7	68	22	AA52368 Human cardiovascular
26	5	41.7	70	22	AA500030 Propionibacterium
27	5	41.7	70	22	AA512788 Human polypeptide
28	5	41.7	72	22	AB503553 Human secreted pro
29	5	41.7	80	22	AA508237 Human musculskele
30	5	41.7	82	23	ABP04929 Human ORFX protein
31	5	41.7	86	22	AA567048 Human immune/haema
32	5	41.7	88	21	AA543048 Human ORFX ORF2812
33	5	41.7	89	22	AA522146 Human cardiovascular
34	5	41.7	90	21	AA508525 Protein encoded by
35	5	41.7	91	22	AA532873 Novel human secret
36	5	41.7	97	22	AA53716 Human EST encoded
37	5	41.7	99	18	AA529891 BRCA-1 fragment 1,
38	5	41.7	99	18	AA522074 BRCA-1 protein kin
39	5	41.7	100	20	AA535267 Chlamydia pneumoni
40	5	41.7	102	21	AA51990 Arabidopsis thalia
41	5	41.7	103	21	AA508049 Propionibacterium
42	5	41.7	105	22	AA542765 Human novel foetal
43	5	41.7	105	22	AA520778 Human polypeptide
44	5	41.7	108	22	AA509641
45	5	41.7	108	22	AA509641

ALIGNMENTS

RESULT 1
AA50916
ID AA50916 standard; Peptide: 12 AA.
XX
AC AA50916;
XX
DT 07-MAY-2002 (first entry)
XX
DE Tumour infiltrating peptide HN-1.
XX
KW Tumour infiltrating peptide; HN-1; head and neck cancer; HNSCC;
KW breast cancer; drug delivery; diagnosis; imaging; gene therapy.
XX
OS Synthetic.
XX
PN WO200202147-A2.
XX
PD 10-JAN-2002.
XX
PE 02-JUL-2001; 2001WO-US21518.
XX
PR 30-JUN-2000; 2000US-215491P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Clayman G, Hong FD;
XX
DR WPI: 2002-195737/25.
XX
PT Peptide internalised by a tumour cell useful for targeted delivery of
PT anticancer drugs -
XX
PS Claim 2; Page 71; 104pp: English.

XX The present sequence is that of a synthetic peptide, termed HN-1,
 CC that is specifically internalised by human head and neck squamous
 CC carcinoma cells (HNSCC), or certain other solid tumour tissue cells,
 CC such as breast cancer cells. HN-1 was identified by screening a
 CC phage M13 peptide library displaying over 10 power 9 peptides. The
 CC screening method was based on the ability of HNSCC line MDAMB27u
 CC cells to uptake peptides by endocytosis at 37 degrees C. The cells
 CC exhibited an approximately 10-fold greater internalisation
 CC potential for HN-1 than normal human fibroblasts. The peptide
 CC localised in cytoplasm after entry. In vivo, i.v. injected HN-1
 CC peptide localised to HNSCC xenograft formed in nude mice. Claimed
 CC compositions comprise HN-1 and a chemotherapeutic, cytotoxic,
 CC apoptotic or DNA-damaging drug, such as taxol. The compositions
 CC are used in claimed methods for killing a tumour cell, especially
 CC an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal
 CC cavity, larynx, thyroid, parathyroid, salivary gland, face or neck
 CC skin or cervical lymph node cell), breast cancer cell or other
 CC solid tumour cell. Also claimed are: methods for detecting cancer
 CC by labelling HN-1 with a detectable label; tumour detection, tumour
 CC imaging and tumour treating kits; methods for killing tumour cells
 CC in which a composition comprising an antitumour compound conjugated
 CC to HN-1 is administered with radiotherapy, chemotherapy, surgery or
 CC a gene therapy composition; and a method for the isolating of an
 CC internalising peptide by phage display library screening. The
 CC peptide provides the necessary dose of a drug specifically to the
 CC tumours, avoiding harmful side effects on other cells. The peptide
 CC is non-toxic, non-immunogenic, stable in vivo, protects its cargo
 CC during transit, and accumulates in a tumour within 48 hours.

XX Sequence 12 AA:

Query Match 100.0%; Score 12; DB 23; Length 12;

Best Local Similarity 100.0%; Pred. No. 3.1e-07;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLNHNGQKL 12

DB 1 TSPLNHNGQKL 12

RESULT 12

ID AAM50917 standard; Peptide: 19 AA.

XX AAM50917:

DT 07-MAY-2002 (first entry)

DE Tumour infiltrating peptide HN-2.

XX Tumour infiltrating peptide; HN-2; head and neck cancer; HNSCC;

KW breast cancer; drug delivery; diagnosis; imaging; gene therapy.

XX Synthetic.

OS Key Location/Qualifiers

FT Peptide 4..15 /note="corresponds to HN-1"

XX WO200202147-A2.

XX 10-JAN-2002.

XX 02-JUL-2001; 2001WO-US21518.

XX 30-JUN-2000; 2000US-215491P.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Clayman G, Hong FD;

XX WPI; 2002-195737/25.

XX Peptide internalised by a tumour cell useful for targeted delivery of
 PT anticancer drugs -
 XX
 PS Example 2; Page 76; 104pp; English.

XX The present sequence is that of a synthetic peptide, termed HN-2,
 CC which is based on the claimed tumour infiltrating peptide HN-1 (see
 CC AAM50916) with additional N- and C-terminal amino acid residues.
 CC HN-1 is specifically internalised by human head and neck squamous
 CC carcinoma cells (HNSCC) and certain other solid tumour tissue cells,
 CC such as breast cancer cells. The additional amino acid residues of
 CC HN-2 did not inhibit cell internalisation of the peptide; HN-1
 CC internalisation is position-independent but sequence-dependent.
 CC Claimed compositions comprise HN-1 and a chemotherapeutic, cytotoxic,
 CC apoptotic or DNA-damaging drug, such as taxol. The compositions
 CC are used in claimed methods for killing a tumour cell, especially
 CC an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal
 CC cavity, larynx, thyroid, parathyroid, salivary gland, face or neck
 CC skin or cervical lymph node cell), breast cancer cell or other
 CC solid tumour cell. Also claimed are: methods for detecting cancer
 CC by labelling HN-1 with a detectable label; tumour detection, tumour
 CC imaging and tumour treating kits; methods for killing tumour cells
 CC in which a composition comprising an antitumour compound conjugated
 CC to HN-1 is administered with radiotherapy, chemotherapy, surgery or
 CC a gene therapy composition; and a method for the isolating of an
 CC internalising peptide by phage display library screening. The
 CC peptide provides the necessary dose of a drug specifically to the
 CC tumours, avoiding harmful side effects on other cells. The peptide
 CC is non-toxic, non-immunogenic, stable in vivo, protects its cargo
 CC during transit, and accumulates in a tumour within 48 hours.

XX Sequence 19 AA:

Query Match 100.0%; Score 12; DB 23; Length 19;

Best Local Similarity 100.0%; Pred. No. 4.8e-07;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLNHNGQKL 12

DB 4 TSPLNHNGQKL 15

RESULT 3

ID AAM50918 standard; Peptide: 19 AA.

XX AAM50918:

DT 07-MAY-2002 (first entry)

DE Tumour infiltrating peptide HN-3.

XX Tumour infiltrating peptide; HN-3; head and neck cancer; HNSCC;

KW breast cancer; drug delivery; diagnosis; imaging; gene therapy.

XX Synthetic.

OS Key Location/Qualifiers

FT Peptide 8..19 /note="corresponds to HN-1"

XX WO200202147-A2.

XX 10-JAN-2002.

XX 02-JUL-2001; 2001WO-US21518.

XX 30-JUN-2000; 2000US-215491P.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Clayman G, Hong FD;

XX WPI; 2002-195737/25.
 DR
 XX
 PT Peptide internalised by a tumour cell useful for targeted delivery of
 anticancer drugs -
 XX
 XX

PS Example 2; Page 76; 104pp; English.

XX The present sequence is that of a synthetic peptide, termed HN-3.
 CC Which is based on the claimed tumour infiltrating peptide HN-1 (see
 CC AAM50916) with additional N-terminal amino acid residues. HN-1 is
 CC specifically internalised by human head and neck squamous carcinoma
 CC cells (HNSCC) and certain other solid tumour tissue cells, such as
 CC breast cancer cells. The additional N-terminal amino acid residues
 CC of HN-3 did not inhibit cell internalisation of the peptide; HN-1
 CC internalisation is position independent but sequence-dependent.
 CC Claimed compositions comprise HN-1 and a chemotherapeutic, cytotoxic,
 CC apoptotic or DNA-damaging drug, such as taxol. The compositions
 CC are used in claimed methods for killing a tumour cell, especially
 CC an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal
 CC cavity, larynx, thyroid, parathyroid, salivary gland, face or neck
 CC skin or cervical lymph node cell), breast cancer cell or other
 CC solid tumour cell. Also claimed are: methods for detecting cancer
 CC by labeling HN-1 with a detectable label; tumour detection, tumour
 CC imaging and tumour treating kits; methods for killing tumour cells
 CC in which a composition comprising an antitumour compound conjugated
 CC to HN-1 is administered with radiotherapy, chemotherapy, surgery or
 CC a gene therapy composition; and a method for the isolating of an
 CC internalising peptide by phage display library screening. The
 CC peptide provides the necessary dose of a drug specifically to the
 CC tumours, avoiding harmful side effects on other cells. The peptide
 CC is non-toxic, non-immunogenic, stable in vivo, protects its cargo
 CC during transit, and accumulates in a tumour within 48 hours.
 XX

SO Sequence 19 AA:

Query Match 100.0%; Score 12; DB 23; Length 19;
 Best Local Similarity 100.0%; Pred. No. 4.8e-07;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPLNHNQOKL 12
 |||||||
 DB 8 TSPLNHNQOKL 19

RESULT 4
 AAY83081

ID AAY83081 standard; Protein: 247 AA.

AC AAY83081;

DT 16-AUG-2000 (first entry)

XX F-box protein FBP-13.

XX F-box protein; FBP; diagnosis; treatment; screening; agonist;
 KM antagonist; proliferative disorder; differentiative disorder;
 KM breast cancer; prostate cancer; ovarian cancer; cancer;
 KM small cell lung carcinoma; immune disorder; cardiovascular disorder;
 KM inflammatory disorder; human.
 XX
 XX

OS Homo sapiens.

PN WO200012679-A1.

PD 09-MAR-2000.

PF 27-AUG-1999; 99WO-US19560.

XX 28-AUG-1998; 98US-009835.
 PR 03-FEB-1999; 99US-0118568.
 PR 15-MAR-1999; 99US-0124449.
 XX

PA (UYNY) UNIV NEW YORK STATE.

XX Chiaur DS, Pagano M, Latres E;

XX WPI; 2000-256635/22.

DR N-PSDB; AA293363.

PT Novel nucleic acid for screening compounds useful for treating
 PT proliferative and differentiative disorders such as cancer and immune
 PT disorders comprises sequences encoding ubiquitin ligases -
 XX
 XX
 PS Claim 10; Figure 16a; 245pp; English.

CC Nucleic acids encoding substrate-targeting subunits of ubiquitin
 CC ligases with F-box motifs (F-box proteins) are useful for diagnosis
 CC of proliferative and differentiated related disorders by measuring
 CC FBP gene expression. Cells expressing such proteins or
 CC their fragments are useful for screening compounds. The compounds
 CC are agonists or antagonists, which are useful for treating a
 CC proliferative or differentiative disorder in a mammal such as
 CC breast, ovarian and prostate cancer and small cell lung carcinoma
 CC and also major opportunistic infections, immune disorders,
 CC cardiovascular diseases and inflammatory disorders. FBP protein,
 CC analogs, derivatives and their subsequences, anti-FBP antibodies
 CC are also useful in diagnosis of the disorders.
 XX

SO Sequence 247 AA:

Query Match 58.3%; Score 7; DB 21; Length 247;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 IHNGOKL 12
 |||||||
 DB 146 IHNGOKL 152

RESULT 5
 AAB48304

ID AAB48304 standard; protein: 408 AA.

AC AAB48304;

DT 02-APR-2001 (first entry)

DE Human ZF23 protein.

XX S-phase kinase associated protein; SKP1; SKP2; SKP2-like protein; ZF;
 KM CUL-1; cullin; CDC53; p27; cyclin E; Max; Mad; c-Myc; MDM2; p53; Bax;
 KM Bad; Bcl-2; tumour; cytostatic.
 XX
 XX

OS Homo sapiens.

PN WO200075184-A1.

PD 14-DEC-2000.

PF 05-JUN-2000; 2000WO-US15449.

PR 04-JUN-1999; 99US-0137494.

PA (UYVA) UNIV YALE.

PI Zhang H, Tsvetkov LM, Kondo T;

DR N-PSDB; AAC84616.

PT Modulating polypeptide levels in a cell, diagnosing and treating tumor,
 PT involves altering levels of proteins such as S-phase kinase associated
 PT proteins 1, 2 and cullin/CDC53 proteins -
 XX
 PS Claim 3; Page 139-140; 162pp; English.

XX The invention relates to methods of altering the polypeptide levels in a
 CC cell, using proteins selected from S-phase kinase associated proteins 1
 CC and 2 (SKP1, SKP2, SKP2-like proteins (ZF) and CUL-1 (a member of the
 CC Cullin/CDC53 family of proteins). The method is useful for altering the
 CC level of p27, cyclin E, Max, Mad, c-Myc, MDM2, p53, Bax, Bad or Bcl-2
 CC polypeptide in a cell. SKP2 and SKP2-like protein levels are useful for
 CC detecting tumours, and in monitoring tumor treatment in a mammal. Agents
 CC that modulate interactions between SKP and target proteins are useful for
 CC treating tumours.

XX Sequence 408 AA:

Query Match 58.3%; Score 7; DB 22; Length 408;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 IHNGOKL 12
 DB 141 IHNGOKL 147

RESULT 5
 ID AAY90287 standard; Protein; 471 AA.

XX AAY90287;

XX 24 OCT-2000 (first entry)

XX Human peptidase, HPEP-4 protein sequence.

XX Human; peptidase; cell proliferative disorder; arteriosclerosis;
 KW psoriasis; myelofibrosis; cancer; autoimmune disorder; Crohn's disease;
 KW inflammatory disorder; AIDS; anaemia; allergy; asthma; atherosclerosis;
 KW Grave's disease; multiple sclerosis; scleroderma; infection; diabetes;
 KW metabolic disorder; Addison's disease; cystic fibrosis; diagnosis;
 KW glycogen storage disease; obesity; therapy; HPEP-4.

XX Homo sapiens.

XX WO200042201-A2.

XX 20-JUL-2000.

XX 11-JAN-2000; 2000WO-US00641.

XX 11-JAN-1999; 99US-0172247.

XX 03-MAY-1999; 99US-0132253.

XX 27-MAY-1999; 99US-0136653.

XX (INCY-) INCYTE PHARM INC.

XX Bandman O, Hillman JL, Tang YT, Azimzai Y, Baughn MR, Lal P;

XX Yue H, Lu DAM;

XX MPI: 2000-482832/42.

XX N-PSDB: AAA37660.

XX An isolated polypeptide for diagnosis, prevention and treatment of

XX cell proliferative, autoimmune/inflammatory and metabolic disorders

XX comprises a sequence encoding a human peptidase -

XX Claim 2: Page 95-96; 131pp; English.

XX This sequence represents a human peptidase, designated HPEP-4. The
 CC invention relates to 18 human peptidases designated HPEP-1 to HPEP-18,
 CC respectively. The peptidases can be used for treating a disease or
 CC condition associated with decreased expression or over expression of
 CC functional human peptidases. The diseases that can be diagnosed,
 CC prevented and treated include cell proliferative disorders (such as
 CC arteriosclerosis, psoriasis, myelofibrosis, and cancers),
 CC autoimmune/inflammatory disorders (such as AIDS, anaemia, allergies,

CC Crohn's disease, asthma, atherosclerosis, Grave's disease, multiple
 CC sclerosis, and scleroderma), infections, and metabolic disorders (such as
 CC Addison's disease, diabetes, cystic fibrosis, glycogen storage diseases
 CC and obesity).

XX Sequence 471 AA:

Query Match 58.3%; Score 7; DB 21; Length 471;
 Best Local Similarity 100.0%; Pred. No. 4.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 IHNGOKL 12
 DB 148 IHNGOKL 154

RESULT 7
 ID AAB93481 standard; Protein; 471 AA.

XX AAB93481;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12772.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELT-) HELIX RES INST.

XX Oka T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX MPI: 2001-318749/34.

XX Claim 8: SEQ ID 12772; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

XX the 5602 nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in

XX the specification. The primer sets can be used in antisense therapy and

XX in gene therapy. The primers are useful for synthesizing polynucleotides,

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 471 AA;

Query Match 58.3%; Score 7; DB 22; Length 471;
 Best Local Similarity 100.0%; Pred. No. 4.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 IHNGOKL 12
 |||||
 DB 148 IHNGOKL 154

RESULT 8
 AAM47764
 ID AAM47764 standard; Protein; 471 AA.

XX
 AC AAM47764;

XX
 DT 26-FEB-2002 (first entry)

XX
 DE F-cassette structural protein 52.

XX
 KM F-cassette structural protein 52; tumour; haemopathy; HIV infection;
 KM immunological disease; inflammation; gene therapy; cytostatic;
 KM haemostatic; virucide; immunomodulatory; antiinflammatory.

XX
 OS Unidentified.

XX
 PN CN1306985-A.

XX
 PD 08-AUG-2001.

XX
 PF 26-JAN-2000; 2000CN-0111548.

XX
 PR 26-JAN-2000; 2000CN-0111548.

XX
 PA (BODA-) BODAO GENE TECH CO LTD SHANGHAI.

XX
 PI Mao Y, Xie Y;

XX
 DR WPI: 2002-026870/04.

XX
 DR N-PSDB; ABA04771.

XX
 PT Polypeptide-F-cassette structural protein 52 and polynucleotide for
 XX coding said polypeptide -

XX
 PS Claim 1; Page 26 (Disclosure); 34pp; Chinese.

XX
 CC The present sequence is the protein sequence for F-cassette structural
 CC protein 52. The protein and its coding sequence are useful in treating
 CC various diseases, such as malignant tumours, haemopathy, HIV infection,
 CC immunological diseases and inflammations.

XX
 SQ Sequence 471 AA;

Query Match 58.3%; Score 7; DB 23; Length 471;
 Best Local Similarity 100.0%; Pred. No. 4.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 IHNGOKL 12
 |||||
 DB 148 IHNGOKL 154

RESULT 9
 ABP38678
 ID ABP38678 standard; Protein; 109 AA.

AC ABP38678;

XX
 DT 24-JUN-2002 (first entry)

XX
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3523.

XX
 KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KM antibacterial; gene therapy.

XX
 OS Staphylococcus epidermidis.

XX
 PN US6380370-B1.

XX
 PD 30-APR-2002.

XX
 PF 13-AUG-1998; 98US-0134001.

XX
 PR 14-AUG-1997; 97US-055779P.

XX
 PR 08-NOV-1997; 97US-064964P.

XX
 PA (GENO-) GENOME THERAPEUTICS CORP.

XX
 PI Doucette-Stamm LA, Bush D;

XX
 DR WPI: 2002-381255/41.

XX
 DR N-PSDB; ABN91223.

XX
 PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 XX Disclosure; SEQ ID 3523; 267pp; English.

XX
 CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life
 CC cycle or inhibit S. epidermidis infection.

XX
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.

XX
 SQ Sequence 109 AA;

Query Match 50.0%; Score 6; DB 23; Length 109;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPUNI 6
 |||||
 DB 40 TSPUNI 45

RESULT 10
 AAG37863
 ID AAG37863 standard; Protein; 130 AA.

XX
 AC AAG37863;

XX
 DT 18-OCT-2000 (first entry)

XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 46628.

XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX
 OS Arabidopsis thaliana.

XX
 PN EP1033405-A2.

PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 9905-0121825.
PR 03-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126284.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 08-APR-1999; 9905-0128234.
PR 16-APR-1999; 9905-0128714.
PR 19-APR-1999; 9905-0129845.
PR 21-APR-1999; 9905-0130077.
PR 23-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 23-APR-1999; 9905-0130831.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.

PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 16-JUL-1999; 9905-0144087.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144336.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0146389.
PR 04-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 06-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 16-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.

PR 23-SEP-1999; 9905-0155486.
 PR 24-SEP-1999; 9905-0155659.
 PR 28-SEP-1999; 9905-0156458.
 PR 29-SEP-1999; 9905-0156596.
 PR 04-OCT-1999; 9905-0157117.
 PR 05-OCT-1999; 9905-0157753.
 PR 06-OCT-1999; 9905-0157865.
 PR 07-OCT-1999; 9905-0158029.
 PR 08-OCT-1999; 9905-0158232.
 PR 12-OCT-1999; 9905-0158369.
 PR 13-OCT-1999; 9905-0158293.
 PR 13-OCT-1999; 9905-0159294.
 PR 13-OCT-1999; 9905-0159295.
 PR 14-OCT-1999; 9905-0159329.
 PR 14-OCT-1999; 9905-0159330.
 PR 14-OCT-1999; 9905-0159331.
 PR 14-OCT-1999; 9905-0159637.
 PR 14-OCT-1999; 9905-0159638.
 PR 18-OCT-1999; 9905-0159584.
 PR 21-OCT-1999; 9905-0160741.
 PR 21-OCT-1999; 9905-0160767.
 PR 21-OCT-1999; 9905-0160768.
 PR 21-OCT-1999; 9905-0160770.
 PR 21-OCT-1999; 9905-0160814.
 PR 21-OCT-1999; 9905-0160815.
 PR 22-OCT-1999; 9905-0160980.
 PR 22-OCT-1999; 9905-0160981.
 PR 22-OCT-1999; 9905-0160989.
 PR 25-OCT-1999; 9905-0161404.
 PR 25-OCT-1999; 9905-0161405.
 PR 25-OCT-1999; 9905-0161406.
 PR 26-OCT-1999; 9905-0161359.
 PR 26-OCT-1999; 9905-0161360.
 PR 26-OCT-1999; 9905-0161361.
 PR 28-OCT-1999; 9905-0161920.
 PR 28-OCT-1999; 9905-0161992.
 PR 28-OCT-1999; 9905-0161993.
 PR 29-OCT-1999; 9905-0162142.

Query Match Best Local Similarity 50.0%; Score 6; DB 21; Length 130;
 Pred. No. 18; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGOK 11
 |||||
 Db 14 IHNGOK 19

RESULT 11
 ABG24051
 ID ABG24051 standard; Protein; 162 AA.
 XX
 AC ABG24051;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #24042.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.

XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR MPI; 2001-639362/73.
 DR N-PSDB; AAS68238.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID NO 54410; 103pp: English.

XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

SEQ Sequence 162 AA;

Query Match Best Local Similarity 50.0%; Score 6; DB 22; Length 162;
 Pred. No. 23; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPINI 6
 |||||
 Db 52 TSPINI 57

RESULT 12
 ABP38390
 ID ABP38390 standard; Protein; 467 AA.
 XX
 AC ABP38390;

DT 24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3235.

KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 antimicrobial; gene therapy.

OS Staphylococcus epidermidis.

PN US6380370-B1.

PD 30-APR-2002.

PF 13-AUG-1998; 98US-0134001.

PR 14-AUG-1997; 97US-055779P.

XX 08-NOV-1997; 97US-064964P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI: 2002-381255/41.
DR N-PSDB: ABN90935.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS Disclosure: SEQ ID 3235; 267pp; English.
XX
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
CC
SQ Sequence 467 AA:

Query Match 50.0%; Score 6; DB 23; Length 467;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 NIHNGQ 10
|||||
DB 456 NIHNGQ 461

RESULT 13

AAG82404
ID AAG82404 standard; Protein: 472 AA.

AC AAG82404;

XX
XX
DT 03-SEP-2001 (first entry)

XX S. epidermidis open reading frame protein sequence SEQ ID NO:1902.

DE
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
KM vaccination; endocarditis.
KM
XX

OS Staphylococcus epidermidis.

XX
XX
PN WO200134809-A2.

XX
XX
PD 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US30782.

XX 09-NOV-1999; 99US-0164258.

XX
XX
PA (GLAXO) GLAXO GROUP LTD.

XX
XX
PI Kimerly WJ;

XX
XX
DR WPI: 2001-316495/33.

XX
XX
DR N-PSDB: AAH53254.

XX
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
useful for vaccinating against infections, e.g. endocarditis -

XX
XX
PS Claim 18; Page 519-520; 2188pp; English.

XX
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4455 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.

SQ Sequence 472 AA:

Query Match 50.0%; Score 6; DB 22; Length 472;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 NIHNGQ 10
|||||
DB 461 NIHNGQ 466

RESULT 14

AAW13187
ID AAW13187 standard; Protein: 20 AA.

AC AAW13187;

XX
XX
DT 19-MAY-1997 (first entry)

XX Fragment of p53 binding protein WBP1.

XX p53; binding protein; WBP1; antibody; diagnosis; treatment;

XX neoplastic; pre-neoplastic; disease; agonist; antagonist;

XX augmentation; inhibition; complex formation; neoplasia; apoptosis;

XX reperfusion; injury; myocardial infarction; stroke; AIDS;

XX traumatic brain; neurodegenerative; aging; ischemia; toxemia;

XX infection; hepatitis; probe; genetic; forensic identification;

XX
XX
OS Homo sapiens.

XX
XX
PN WO9514777-A1.

XX
XX
PD 01-JUN-1995.

XX 21-NOV-1994; 94WO-US13499.

XX 22-NOV-1993; 93US-0156571.

XX
XX
PA (ONYX-) ONYX PHARM.

XX
XX
PI Bischoff JR, Wu L;

XX
XX
DR WPI: 1995-206934/27.

XX
XX
PT New p53-binding polypeptide(s) WBP1 and p53BNC - used to develop
PT prods. for screening assays and for use in diagnosis and therapy of
PT diseases, esp. neoplasia

XX
XX
PS Claim 8; Page 38; 90pp; English.

XX
XX
CC The present sequence is a fragment of the p53 binding protein WBP1,
CC which can be used to generate antibodies for the diagnosis of
CC (pre)neoplastic diseases, and WBP1 (ant)agonists. The (ant)agonists
CC augment or inhibit the formation of p53:WBP1 complexes and
CC inhibit neoplasia or apoptosis, useful in the treatment of, e.g.
CC reperfusion injury, myocardial infarction, stroke, traumatic brain
CC injury, neurodegenerative diseases, aging, ischemia, toxemia,

THIS PAGE BLANK (USPTO)

CC infection, AIDS and hepatitis. The WP1 encoding cDNA, which was
CC isolated from a HeLa cell derived cDNA library using the yeast
CC two-hybrid system, can be used to generate probes for the diagnosis
CC of (pre)neoplastic pathological conditions and genetic diseases,
CC and the forensic identification of human individuals.

XX
SQ Sequence 20 AA;

Query Match 41.7%; Score 5; DB 16; Length 20;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPIN 5
11111
Db 15 TSPIN 19

RESULT 15

AAW57358
ID AAW57358 standard; peptide: 20 AA.

XX
AC AAW57358;

XX
DT 11-AUG-1998 (first entry)

DE Human WP1 immunogenic peptide SEQ ID NO:90.

XX
KM Human: WP1; p53; cancer; interacting protein; screen; diagnosis;

KW genetic disease; forensic identification; nuclear phosphoprotein;
KW cellular proliferation; neoplastic transformation; p53UBC; Immunogen.

XX
OS Homo sapiens.

XX
PN US5756669-A.

XX
PD 26-MAY-1998.

XX
PF 02-MAR-1995; 95US-0399696.

XX
PR 02-MAR-1995; 95US-0399696.

XX
PR 23-NOV-1993; 93US-0156571.

XX
PR 21-NOV-1994; 94WO-US13499.

XX
PA (ONYX-) ONYX PHARM INC.

XX
PI Bischoff JR, Wu L;

XX
DR WP1: 1998-321626/28.

XX
PT WP1, p53-interacting protein - useful screening agents for
XX treatments of p53 related cancers

XX
PS Disclosure; Column 24; 68pp; English.

XX
CC The present sequence represents an immunogenic peptide from WP1
CC (cellular proliferation and neoplastic transformation) polypeptide,
CC which can be used to screen bacteriophage antibody display libraries or
CC to immunize a rabbit. WP1 is a p53-interacting protein that can be used
CC to screen for agents for diagnosis or treatment of cancer and genetic
CC diseases and for forensic identification of human individuals. The p53
CC protein is a nuclear phosphoprotein involved in control of cellular
CC proliferation, and mutations in it are associated with human cancers.

XX
SQ Sequence 20 AA;

Query Match 41.7%; Score 5; DB 19; Length 20;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TSPIN 5
11111
Db 15 TSPIN 19

Search completed: January 3, 2003, 14:27:45
Job time : 37 secs